

ANNA BARAŃSKA

ELEMENTY PROBABILISTYKI
I STATYSTYKI MATEMATYCZNEJ
W INŻYNIERII ŚRODOWISKA

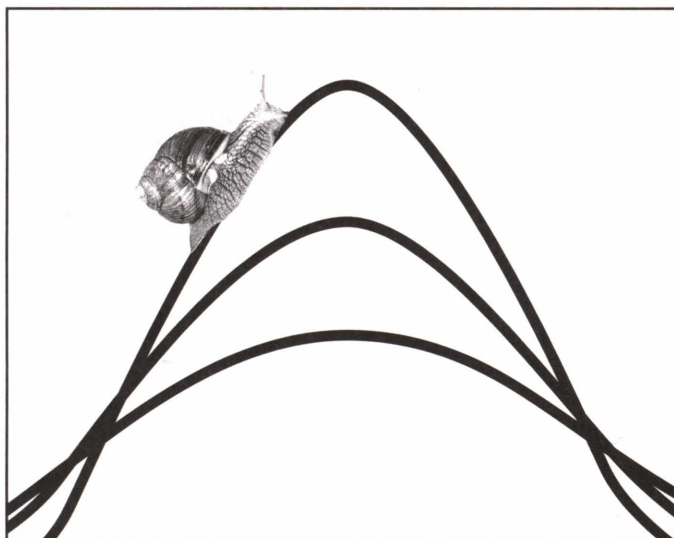


UCZELNIANE WYDAWNICTWA NAUKOWO-DYDAKTYCZNE

KRAKÓW 2008

ANNA BARAŃSKA

ELEMENTY PROBABILISTYKI
I STATYSTYKI MATEMATYCZNEJ
W INŻYNIERII ŚRODOWISKA



UCZELNIANE WYDAWNICTWA NAUKOWO-DYDAKTYCZNE

KRAKÓW 2008

KU 0252 pozycja wydawnictw naukowych
Akademii Górniczo-Hutniczej im. S. Staszica w Krakowie

© Wydawnictwa AGH, Kraków 2008
ISBN 978-83-7464-118-0

Redaktor Naczelny Uczelnianych Wydawnictw
Naukowo-Dydaktycznych AGH: *Jan Sas*

Z-ca Redaktora Naczelnego: *Beata Barszczewska-Wojda*

Komitety Naukowy UWND AGH:
Kazimierz Jeleń (przewodniczący),
Edward Fraś,
Tadeusz Sawik,
Ryszard Uberman,
Adam Paweł Wojda,
Mariusz Ziółko



Recenzent: *dr hab. inż. Edward Preweda, prof. AGH*

Redakcja: *Małgorzata Koch*

Projekt okładki, strony tytułowej i redakcja techniczna: *Beata Barszczewska-Wojda*

Opracowanie graficzne rysunków: *dr inż. Marek Kulczycki*

Skład komputerowy: **LIBRA**, tel. 0 600 053 872

Redakcja Uczelnianych Wydawnictw Naukowo-Dydaktycznych AGH
al. Mickiewicza 30, 30-059 Kraków
tel. 012 617 32 28, tel./faks 012 636 40 38
e-mail: redakcja@wydawnictwoagh.pl
<http://www.WydawnictwoAGH.pl>

Spis treści

Wykaz symboli	5
Wstęp	9
I. Elementy probabilistyki	11
I.1. Podstawowe pojęcia rachunku prawdopodobieństwa	11
I.1.1. Przestrzeń probabilistyczna	11
I.1.2. Prawdopodobieństwo warunkowe i niezależność zdarzeń	12
I.2. Zmienna losowa i dystrybuenta	15
I.2.1. Jednowymiarowa zmienna losowa dyskretna	15
I.2.2. Jednowymiarowa zmienna losowa ciągła	19
I.2.3. Charakterystyki liczbowe rozkładu zmiennej losowej	21
I.2.4. Wybrane rozkłady zmiennej losowej dyskretnej	25
I.2.5. Wybrane rozkłady zmiennej losowej ciągłej	31
I.3. Zmienna losowa dwuwymiarowa	43
I.3.1. Dwuwymiarowa zmienna losowa dyskretna	43
I.3.1.1. Rozkłady brzegowe zmiennych losowych dyskretnych	45
I.3.1.2. Rozkłady warunkowe zmiennych losowych dyskretnych	46
I.3.2. Dwuwymiarowa zmienna losowa ciągła	47
I.3.2.1. Rozkłady brzegowe zmiennych losowych ciągłych	48
I.3.2.2. Rozkłady warunkowe zmiennych losowych ciągłych	49
I.3.3. Niezależność zmiennych losowych	49
I.3.4. Charakterystyki liczbowe zmiennej losowej dwuwymiarowej	50
I.3.5. Współczynnik korelacji	54
I.3.6. Proste regresje drugiego rodzaju	55
I.3.7. Dwuwymiarowy rozkład normalny	58
II. Elementy statystyki matematycznej	61
II.1. Populacja, próba, cecha	61
II.2. Statystyka opisowa	63
II.2.1. Szereg rozdzielczy	63
II.2.2. Wartość średnia i miary rozproszenia	65

II.3. Wybrane rozkłady z próby.....	68
II.3.1. Rozkład średniej arytmetycznej z próby	68
II.3.2. Rozkład wariancji z próby	70
II.3.3. Rozkład ilorazu wariancji z prób	72
II.3.4. Rozkład ilorazu wartości przeciętnej i odchylenia standardowego z próby	74
II.3.5. Rozkład frakcji w próbce.....	75
II.4. Estymacja.....	76
II.4.1. Własności estymatorów	76
II.4.2. Metody estymacji	77
II.4.2.1. Metoda najmniejszych kwadratów	77
II.4.2.2. Metoda największej wiarygodności.....	80
II.4.3. Estymacja punktowa.....	82
II.4.4. Estymacja przedziałowa	84
II.4.4.1. Estymacja przedziałowa wartości przeciętnej	84
II.4.4.2. Estymacja przedziałowa wariancji i odchylenia standardowego	88
II.4.4.3. Estymacja przedziałowa dla wskaźnika struktury	92
II.5. Testowanie hipotez statystycznych	94
II.5.1. Testy parametryczne.....	95
II.5.1.1. Testowanie pojedynczej wartości przeciętnej	95
II.5.1.2. Testowanie pojedynczej wariancji.....	101
II.5.1.3. Test porównujący dwie wariancje.....	104
II.5.1.4. Testy porównujące dwie wartości przeciętne.....	105
II.5.1.5. Testowanie wartości wskaźnika struktury	108
II.5.1.6. Testy porównujące dwa wskaźniki struktury	110
II.5.1.7. Konsekwencje niespełnienia założenia o normalności rozkładu	112
II.5.2. Testy nieparametryczne.....	112
II.5.2.1. Test χ^2 -Pearsona.....	113
II.5.2.2. Test Kołmogorowa	118
II.6. Regresja	125
Dodatek	
Zasady korzystania z tablic statystycznych	129
Tablice statystyczne	133
Literatura	163
Skorowidz	165

Wykaz symboli

ω	–	zdarzenie elementarne
Ω	–	przestrzeń zdarzeń elementarnych
Σ	–	sigma-algebra
A, B	–	zbiory, zdarzenia
A', B'	–	dopełnienia zbiorów, zdarzenia przeciwne
$\#A$	–	moc (liczność) zbioru A
\emptyset	–	zbiór pusty, zdarzenie niemożliwe
$P(A)$	–	prawdopodobieństwo zdarzenia A
p_i	–	prawdopodobieństwo w rozkładzie brzegowym
p	–	wskaźnik struktury (frakcja), poziom ufności
α	–	poziom istotności
$A B$	–	zdarzenie A pod warunkiem zdarzenia B
\forall_i	–	dla każdego i
\exists_i	–	istnieje i
$\bigcup_i A_i$	–	suma mnogościowa po i zdarzeń A_i
$\sum_i x_i$	–	suma algebraiczna po i liczb x_i
$\prod_i x_i$	–	iloczyn algebraiczny po i liczb x_i
$\lim_{x \rightarrow x_0^+} F(x)$	–	prawostronna granica funkcji F w punkcie x_0
$\lim_{x \rightarrow x_0^-} F(x)$	–	lewostronna granica funkcji F w punkcie x_0
X, Y, Z, T, F	–	zmienne losowe
F	–	zmienna losowa o rozkładzie F-Snedecora
T	–	zmienna losowa o rozkładzie T-Studenta

- χ^2 – zmienna losowa o rozkładzie chi-kwadrat
- $|X|$ – wartość bezwzględna zmiennej X
- δ – odchyłka
- μ – wartość oczekiwana w rozkładzie normalnym
- $X \sim N(\mu, \sigma)$ – zmienna X ma rozkład normalny o parametrach μ, σ
- x_i^+ – następna w kolejności wartość zmiennej po x_i
- p_i – prawdopodobieństwo zdarzenia (zmiennej) o numerze i
- f – funkcja gęstości
- $f(x|y)$ – funkcja gęstości w rozkładzie warunkowym
- $F(x)$ – wartość funkcji dystrybuanty
- F' – pochodna z funkcji F
- $\frac{\partial f(x, y)}{\partial x}$ – pochodna cząstkowa z funkcji f po zmiennej x
- $F(x, y)$ – dystrybuanta dwuwymiarowa w punkcie (x, y)
- φ – gęstość standaryzowanego rozkładu normalnego
- Φ – dystrybuanta standaryzowanego rozkładu normalnego
- Γ – funkcja gamma
- $z(\alpha)$ – kwantyl standaryzowanego rozkładu normalnego dla prawdopodobieństwa α
- $t(\alpha, k)$ – kwantyl rozkładu T-Studenta dla prawdopodobieństwa α i k stopni swobody
- $\chi^2(\alpha, k)$ – kwantyl rozkładu chi-kwadrat dla prawdopodobieństwa α i k stopni swobody
- $\exp(x)$ – wartość funkcji wykładniczej e^x
- N^+ – liczby naturalne dodatnie (bez zera)
- R^2 – zbiór punktów na płaszczyźnie o współrzędnych rzeczywistych
- γ – współczynnik asymetrii (skośność)
- K – współczynnik skupienia (kurtoza)
- E – współczynnik spłaszczenia (eksces)
- \hat{X} – estymator wartości średniej
- \hat{x} – średnia z wartości x_i
- $E(X)$ – wartość oczekiwana zmiennej X
- $E(\hat{X})$ – wartość oczekiwana estymatora wartości średniej
- $V(X)$ – wariancja zmiennej X
- $V(\hat{X})$ – wariancja estymatora wartości średniej
- \bar{S}^2 – obciążony estymator wariancji z próby losowej

- \hat{S}^2 – nieobciążony estymator wariancji z próby losowej
- S^2 – nieobciążony estymator wariancji z próby o znanej wartości przeciętnej
- $E(\hat{S}^2)$ – wartość oczekiwana estymatora wariancji
- $V(\hat{S}^2)$ – wariancja estymatora wariancji
- $\sigma(X)$ – odchylenie standardowe zmiennej X
- $cov(X, Y)$ – kowariancja zmiennych X i Y
- $Cov(X, Y)$ – macierz kowariancji zmiennych X i Y
- $Cov^{-1}(X, Y)$ – odwrotność macierzy kowariancji
- $\det(Cov(X, Y))$ – wyznacznik macierzy kowariancji
- a_r – moment zwykły absolutny rzędu r
- b_r – moment centralny absolutny rzędu r
- $\alpha_r(X)$ – moment zwykły rzędu r zmiennej X
- α_{rs} – moment zwykły mieszany rzędu r, s
- $\mu_r(X)$ – moment centralny rzędu r zmiennej X
- μ_{rs} – moment centralny mieszany rzędu r, s
- β – kąt skręcenia elipsy
- q – współczynnik korelacji krzywoliniowej
- r – współczynnik korelacji liniowej
- H_0 – hipoteza zerowa
- H_1 – hipoteza alternatywna

Wstęp

Książka składa się z dwóch zasadniczych działów. Pierwszy z nich traktuje o wybranych zagadnieniach z zakresu rachunku prawdopodobieństwa. Jest w zasadzie wprowadzeniem do działu drugiego, omawiającego podstawowe metody analiz statystycznych i stanowiącego zasadniczą część podręcznika. Bazuje się tutaj w dużym stopniu na wiedzy studentów wyniesionej jeszcze ze szkoły oraz z kursu matematyki na pierwszych semestrach studiów, między innymi na wiedzy z zakresu podstaw rachunku macierzowego, który został pominięty w niniejszej pozycji.

Z zakresu probabilistyki omówiono podstawowe pojęcia, koncentrując się na zmiennej losowej, będącej punktem wyjścia zapisu danych pomiarowych w formie nadającej się do analiz statystycznych. Scharakteryzowane zostały przykładowe rozkłady prawdopodobieństwa zmiennych losowych dyskretnych oraz ciągłych, a także podstawowe charakterystyki opisujące postać rozkładu i zależności pomiędzy parą zmiennych, ze szczególnym uwzględnieniem pojęcia regresji liniowej. Początek działu drugiego (rozdziały II.1–II.3) to teoretyczny wstęp do omówienia wybranych metod analiz statystycznych, z których bardziej szczegółowo zatrzymano się nad zagadnieniami estymacji punktowej i przedziałowej oraz weryfikacji hipotez za pomocą wybranych parametrycznych i nieparametrycznych testów istotności.

Liczę na to, że niniejszy podręcznik stanie się pomocny studentom w przyswajaniu materiału napiętego programu. Został on napisany dość wcześnie, biorąc pod uwagę mój krótki staż pracy w charakterze wykładowcy oraz fakt, że miałam okazję prowadzić wykład z tego zakresu zaledwie przez trzy lata. Nieuniknioną konsekwencją tych okoliczności jest niedoskonałość niniejszego podręcznika. Skłania mnie to do przyjęcia otwartej postawy wobec wszelkich uwag ze strony Czytelników. Mam nadzieję, że uwagi te przyczynią się do udoskonalenia ewentualnych ponownych edycji podręcznika. Treść podręcznika w niewielkim stopniu wykracza poza program wykładów i ćwiczeń, więc z pewnością nie będzie wystarczająca dla osób zainteresowanych poszerzeniem wiedzy w tej dziedzinie, których odsyłam do pozycji literatury zamieszczonych na końcu książki. Podczas pisania niniejszego podręcznika sama, w dużej mierze, na tej literaturze się opierałam, starając się trudno brzmiące sformułowania wyrazić możliwie najprościej – zachowując ich naukowy charakter – by łatwiej je było zrozumieć. Czytelnicy zweryfikują, w jakim stopniu to się powiodło. Również niektóre przykłady zostały zaadaptowane do niniejszego podręcznika z innych pozycji literatury.

Mam również nadzieję, że książka pomoże przezwyciężyć obawę przed metodami statystycznymi i przełamać – tak częsty wśród studentów – stereotyp zakładający, iż są one zbyt trudne, by je zrozumieć. A raczej, choć w niewielkim stopniu, przyczyni się do ukazania uniwersalnych możliwości ich zastosowania w zagadnieniach różnych dziedzin, których przedmiotem jest analiza konkretnych danych pomiarowych.

Z uwagi na bezsporną powszechność zastosowania metod statystycznych, zarówno w naukach przyrodniczych, technicznych, jak i humanistycznych, niniejszy podręcznik ma szansę okazać się przydatny dla studentów do ich dalszej nauki i pracy.

I. Elementy probabilistyki

I.1. Podstawowe pojęcia rachunku prawdopodobieństwa

I.1.1. Przestrzeń probabilistyczna

Każda współczesna teoria matematyczna jest teorią aksjomatyczną, czyli taką, w której wszystkie twierdzenia wyprowadza się z aksjomatów tej teorii.

Aksjomat to zdanie, którego prawdziwość przyjmuje się bez dowodu, formułujące zależności między tzw. pojęciami pierwotnymi danej teorii. Nie inaczej jest z teorią rachunku prawdopodobieństwa. Ma ona swoje **pojęcia pierwotne**, czyli takie, których się nie definiuje. Należą do nich:

- zdarzenie elementarne,
- przestrzeń zdarzeń elementarnych.

Przestrzeń zdarzeń elementarnych nazywa się również **zbiorem zdarzeń elementarnych** i oznacza się najczęściej przez Ω .

Zdarzenia elementarne, oznaczane ω , są to elementy przestrzeni Ω . W praktyce, przestrzeń zdarzeń elementarnych to zbiór wszystkich możliwych, niepodzielnych wyników obserwacji lub doświadczenia.

Definicja I.1.1. Rodzinę podzbiorów Σ , przestrzeni zdarzeń elementarnych Ω , spełniającą następujące warunki:

- (1) $\Omega \in \Sigma$,
- (2) $A \in \Sigma \Rightarrow \Omega \setminus A \in \Sigma$,
- (3) $A_1, A_2, A_3, \dots \in \Sigma \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \Sigma$.

nazywamy **sigma-algebrą** (σ -algebrą). Elementy sigma-algebry Σ to tzw. **zdarzenia losowe**. Należą do nich cała przestrzeń Ω , zwana **zdarzeniem pewnym**, oraz jej dopełnienie, czyli zbiór pusty \emptyset , zwany **zdarzeniem niemożliwym**.

Prawe strony implikacji w warunkach (2) i (3) oznaczają wyniki działań na zdarzeniach, odpowiednio różnicy i sumy, które odpowiadają analogicznym działaniom na zbiorach. Zatem alternatywa zdarzeń $A \vee B$ odpowiada sumie mnogościowej zbiorów $A \cup B$, koniunkcja

zdarzeń $A \wedge B$ – iloczynowi mnogościowemu zbiorów $A \cap B$, a różnica zdarzeń $A \setminus B$ jest określana tak samo jak różnica w teorii zbiorów. Różnica $\Omega \setminus A$, odpowiada dopełnieniu zbioru A w przestrzeni Ω , a w teorii prawdopodobieństwa jest to zdarzenie przeciwne do zdarzenia A , oznaczane A' .

Po takim wstępie, wyjaśniającym podstawowe pojęcia rachunku prawdopodobieństwa, możemy podać definicję przestrzeni probabilistycznej.

Definicja I.1.2. Przestrzeń probabilistyczną nazywamy trójkę (Ω, Σ, P) , gdzie P jest miarą unormowaną, czyli spełniającą następujące warunki:

- (1) $A \in \Sigma \Rightarrow P(A) \geq 0$;
- (2) zbiory $A_1, A_2, A_3, \dots \in \Sigma$ są parami rozłączne $\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$;
- (3) $P(\Omega) = 1$.

P nazywane jest **miarą probabilistyczną**. Jest to funkcja określona na σ -algebrze, o wartościach w przedziale $[0; 1]$, co symbolicznie zapisujemy $P: \Sigma \rightarrow [0; 1]$. Wartość funkcji P dla zdarzenia A jest **prawdopodobieństwem** tego zdarzenia.

Z definicji I.1.2 wynikają następujące własności prawdopodobieństwa:

- (1) $P(\emptyset) = 0$
- (2) $A \subset B \Rightarrow P(A) \leq P(B)$
- (3) $A \in \Sigma \Rightarrow P(A) \leq 1$
- (4) $A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$ (I.1.1)
- (5) $P(A) + P(A') = 1$
- (6) $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- (7) jeżeli przestrzeń Ω składa się z n zdarzeń elementarnych ω_i , czyli $\#\Omega = n$, oraz zdarzenia elementarne są jednakowo prawdopodobne, czyli $P(\omega_i) = 1/n$, to prawdopodobieństwo zdarzenia A składającego się z k zdarzeń elementarnych wynosi:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{k}{n},$$

gdzie: $\#A$, $\#\Omega$ oznacza tzw. moc (liczność) zbioru, odpowiednio A oraz Ω .

Ostatnia z powyższych własności to tzw. klasyczna definicja prawdopodobieństwa.

I.1.2. Prawdopodobieństwo warunkowe i niezależność zdarzeń

Definicja I.1.3. Prawdopodobieństwo warunkowe to prawdopodobieństwo zajścia zdarzenia A przy założeniu, że zaszło zdarzenie B . Jeżeli A i B należą do sigma-algebry Σ w tej

samej przestrzeni probabilistycznej (Ω, Σ, P) oraz $B \neq \emptyset$, to prawdopodobieństwo warunkowe $P(A|B)$ określa się następująco:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{gdzie } A, B \in \Sigma, \quad P(B) > 0 \quad (\text{I.1.2})$$

Znając pojęcie prawdopodobieństwa warunkowego, można sformułować intuicyjny warunek niezależności zdarzeń.

Definicja I.1.4. A i B są **zdarzeniami niezależnymi**, jeżeli prawdopodobieństwo każdego z nich nie zależy od warunku zadanego przez drugie zdarzenie, tzn.:

$$P(A|B) = P(A) \quad \text{oraz} \quad P(B|A) = P(B) \quad (\text{I.1.3})$$

Podstawiając do lewych stron obu równań (I.1.3) wyrażenie (I.1.2), uzyskujemy warunek niezależności zdarzeń:

$$P(A \cap B) = P(A) \cdot P(B) \quad (\text{I.1.4})$$

Analogicznie formułujemy definicję dla dowolnej, skończonej liczby n zdarzeń:

Definicja I.1.5. Zdarzenia A_1, A_2, \dots, A_n są niezależne, jeżeli dla każdego ich podzbioru r -elementowego zachodzi równość:

$$P(A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_r}) = P(A_{k_1}) \cdot P(A_{k_2}) \cdot \dots \cdot P(A_{k_r}) \quad (\text{I.1.5})$$

Z prawdopodobieństwami warunkowymi spotykamy się najczęściej przy obliczaniu tzw. prawdopodobieństwa całkowitego. Mówi o nim następujące twierdzenie:

Twierdzenie I.1.1 (o prawdopodobieństwie całkowitym)

ZAŁOŻENIA

(Ω, Σ, P) – przestrzeń probabilistyczna,

A_1, A_2, \dots, A_n – zdarzenia należące do Σ , takie, że:

$$(1) \quad \forall_{i=1, 2, \dots, n} P(A_i) > 0,$$

$$(2) \quad \forall_{i \neq j} A_i \cap A_j = \emptyset,$$

$$(3) \quad \bigcup_{i=1}^n A_i = \Omega.$$

TEZA

$$\forall_{B \in \Sigma} P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i).$$

Dowód

$$\text{Z założeń mamy: } B = B \wedge \Omega = B \wedge \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (B \wedge A_i).$$

$$\text{Zatem: } P(B) = \sum_{i=1}^n P(B \wedge A_i) = \sum_{i=1}^n P(B | A_i) \cdot P(A_i).$$

c.b.d.u.

Z twierdzeniem I.1.1 wiąże się ściśle następujące twierdzenie:

Twierdzenie I.1.2. (Bayesa)

ZAŁOŻENIA

(Ω, Σ, P) – przestrzeń probabilistyczna,

A_1, A_2, \dots, A_n – zdarzenia należące do Σ , takie, że:

$$(1) \quad \forall_{i=1, 2, \dots, n} P(A_i) > 0,$$

$$(2) \quad \forall_{i \neq j} A_i \wedge A_j = \emptyset,$$

$$(3) \quad \bigcup_{i=1}^n A_i = \Omega.$$

TEZA

$$\forall_{B \in \Sigma} \quad P(A_k | B) = \frac{P(B | A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B | A_i) \cdot P(A_i)}.$$

Dowód

Z definicji prawdopodobieństwa warunkowego I.1.3. i z tezy wcześniejszego twierdzenia I.1.1. mamy:

$$P(A_k | B) = \frac{P(B \wedge A_k)}{P(B)} = \frac{P(B | A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B | A_i) \cdot P(A_i)}.$$

c.b.d.u.

Twierdzenie Bayesa można parafrazować następująco: jeżeli skutek B nastąpił w wyniku jednej z przyczyn A_1, A_2, \dots, A_n , jedynych możliwych i wzajemnie wykluczających się, to prawdopodobieństwo tego, że A_k ($k = 1, 2, \dots, n$) było przyczyną zajścia tego skutku, wynosi $P(A_k | B)$.

I.2. Zmienna losowa i dystrybuanta

Zmienna losowa, jako liczbową reprezentacją wyniku doświadczenia losowego, jest kluczowym pojęciem analiz statystycznych. Pojęcie losowości jest ściśle związane z faktem niemożliwości dokładnego przewidzenia wyniku doświadczenia.

Zacznijmy od przytoczenia definicji zmiennej losowej i dystrybuanty.

Definicja I.2.1. Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną. **Zmienna losowa** X to funkcja określona na przestrzeni zdarzeń elementarnych Ω , o wartościach w zbiorze liczb rzeczywistych R , czyli $X: \Omega \rightarrow R$, spełniająca następujący warunek: zbiór zdarzeń elementarnych, dla których wartości zmiennej losowej X są mniejsze od x , jest zdarzeniem. Symbolicznie możemy to zapisać:

$$\forall_{x \in R} \{ \omega: X(\omega) < x \} \in \Sigma \quad (\text{I.2.1})$$

Dla skończonej przestrzeni Ω , w której zdarzeniami są wszystkie jej podzbiory, dowolna funkcja $X: \Omega \rightarrow R$ jest zmienną losową. Zmienne losowe oznaczamy dużymi literami, np.: X , Y , Z , natomiast ich wartości małymi x , y , z .

Definicja I.2.2. Funkcja $F: R \rightarrow [0; 1]$ jest **dystrybuantą zmiennej losowej** X , jeżeli:

$$\forall_{x_0 \in R} F(x_0) = P(X < x_0).$$

Własności dystrybuanty:

- (1) $\forall_{x \in R} F(x) \in [0; 1]$
 - (2) $\lim_{x \rightarrow -\infty} F(x) = 0$ oraz $\lim_{x \rightarrow +\infty} F(x) = 1$
 - (3) F jest funkcją niemalejącą, czyli $\forall_{x_1, x_2 \in R} x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
 - (4) F jest lewostronnie ciągła, czyli $\forall_{x_0 \in R} \lim_{x \rightarrow x_0^-} F(x) = F(x_0)$
 - (5) $P(a \leq X < b) = F(b) - F(a)$
 - (6) $P(X = x_0) = \lim_{x \rightarrow x_0^+} F(x) - F(x_0)$
- (I.2.2)

I.2.1. Jednowymiarowa zmienna losowa dyskretna

Definicja I.2.3. Jeżeli zmienna losowa X ma co najwyżej przeliczalny zbiór wartości $\{x_1, x_2, x_3, \dots\}$ taki, że:

$$P(X = x_i) = p_i > 0, \quad i \in N, \quad \sum_i p_i = 1,$$

to nazywamy ją **zmienną losową dyskretną** (skokową).

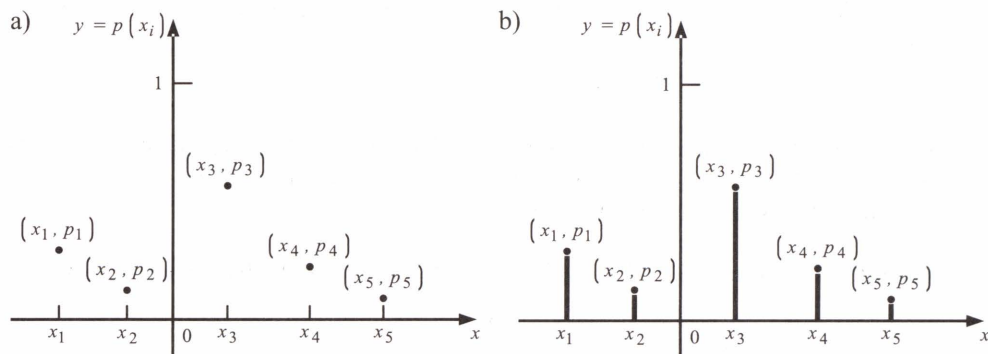
Wartości x_i wraz z odpowiadającymi im wartościami prawdopodobieństw p_i tworzą tzw. **rozkład prawdopodobieństwa zmiennej losowej** X , najczęściej zapisywany w postaci tabeli I.2.1.

Tabela I.2.1

Rozkład prawdopodobieństwa zmiennej losowej dyskretnej X

x_i	x_1	x_2	x_3	...
p_i	p_1	p_2	p_3	...

Punkty o współrzędnych (x_i, p_i) umieszczone w układzie współrzędnych tworzą wykres tzw. funkcji prawdopodobieństwa (rys. I.2.1a). Łącząc te punkty prostokątami z osią OX , uzyskujemy tzw. histogram rozkładu prawdopodobieństwa (rys. I.2.1b). Oczywiście suma długości wszystkich odcinków wynosi 1.



Rys. I.2.1. Wykres funkcji rozkładu prawdopodobieństwa (a); histogram funkcji rozkładu prawdopodobieństwa (b)

Definicja I.2.2. dystrybucyjny zmiennej losowej w przypadku zmiennej skokowej przyjmuje postać:

$$F(x_0) = P(X < x_0) = \sum_{x_i < x_0} p_i \quad (\text{I.2.3})$$

Pomiędzy dystrybucyjną i rozkładem zmiennej losowej istnieje odpowiedniość wzajemnie jednoznaczna. A zatem podobnie jak wyznaczamy wartości funkcji dystrybucyjnej z rozkładu prawdopodobieństwa, tak również możemy wyznaczyć rozkład, znając wartości dystrybucyjnej. Korzystamy przy tym z własności (6) dystrybucyjnej (I.2.2).

PRZYKŁAD I.2.1

Sprawdzić, czy dane (tab. I.2.2) mogą być rozkładem zmiennej losowej dyskretnej Y .

Tabela I.2.2

Wartości zmiennej losowej Y wraz z prawdopodobieństwami ich występowania

y_i	-5	-2	1	4	6	7
p_i	0,05	0,10	0,35	0,25	0,20	0,05

Jeśli tak, podać wartości dystrybuanty dla tej zmiennej, narysować jej wykres, obliczyć podane prawdopodobieństwa i zapisać je za pomocą funkcji dystrybuanty $F(y)$.

- a) $P(Y \geq 3)$,
- b) $P(-5 < Y < 5)$,
- c) $P(Y > 6)$.

Rozwiązanie

W celu sprawdzenia, czy mamy do czynienia z rozkładem prawdopodobieństwa, zgodnie z definicją I.2.3, sumujemy wszystkie prawdopodobieństwa. Suma wynosi 1, zatem rzeczywiście dane z tabeli I.2.2 mogą być rozkładem zmiennej losowej Y . Wartości dystrybuanty wyznaczamy zgodnie z jej definicją, ze wzoru (I.2.3):

$$F(-5) = P(Y < -5) = \sum_{y_i < -5} p_i = 0,$$

$$F(-2) = P(Y < -2) = \sum_{y_i < -2} p_i = 0,05,$$

$$F(1) = P(Y < 1) = \sum_{y_i < 1} p_i = 0,05 + 0,10 = 0,15 \text{ itd.}$$

Pamiętać należy, że zgodnie z definicją I.2.2, dystrybuita określona jest na całym zbiorze liczb rzeczywistych. Dlatego do narysowania wykresu potrzebne są nam również jej wartości dla liczb spoza przestrzeni zdarzeń elementarnych Ω . Wyznamy kilka z nich:

$$F(-10) = P(Y < -10) = \sum_{y_i < -10} p_i = 0,$$

$$F(4,5) = P(Y < 4,5) = \sum_{y_i < 4,5} p_i = 0,05 + 0,10 + 0,35 + 0,25 = 0,75,$$

$$F(11) = P(Y < 11) = \sum_{y_i < 11} p_i = 0,05 + 0,10 + 0,35 + 0,25 + 0,20 + 0,05 = 1,00.$$

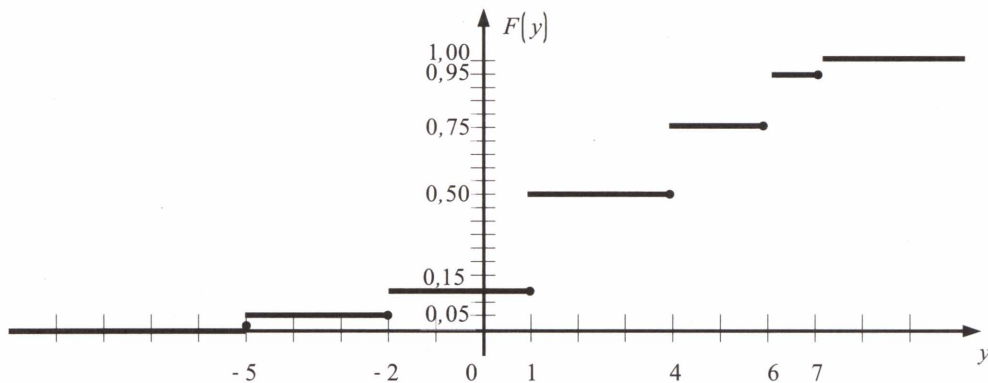
W tabeli I.2.3 zamieszczono obliczone wartości dystrybuanty dla wartości zmiennej losowej Y .

Tabela I.2.3
Wartości funkcji dystrybuanty zmiennej losowej Y

y_i	-5	-2	1	4	6	7	$Y > 7$
$F(y_i)$	0,00	0,05	0,15	0,50	0,75	0,95	1,00

Wykres funkcji dystrybuanty przedstawiony jest na rysunku I.2.2.





Rys. I.2.2. Wykres funkcji dystrybuanty

Pozostaje obliczyć trzy prawdopodobieństwa i wyrazić je za pomocą wartości dystrybuanty:

$$\text{a) } P(Y \geq 3) = P(Y = 4 \vee Y = 6 \vee Y = 7) = 0,25 + 0,20 + 0,05 = 0,50,$$

z drugiej strony mamy:

$$P(Y \geq 3) = 1 - P(Y < 3) = 1 - P(Y < 4) = 1 - F(4) = 1 - 0,50 = 0,50;$$

$$\text{b) } P(-5 < Y < 5) = P(Y = -2 \vee Y = 1 \vee Y = 4) = 0,10 + 0,35 + 0,25 = 0,70,$$

$$P(-5 < Y < 5) = P(Y < 5) - P(Y \leq -5) = F(5) - F(-5^+) = F(6) - F(-2) = 0,75 - 0,05 = 0,70,$$

przy czym $F(-5^+)$ jest umownym zapisem dla oznaczenia wartości dystrybuanty w punkcie następnym w stosunku do -5 w rozkładzie zmiennej losowej;

$$\text{c) } P(Y > 6) = P(Y = 7) = 0,05;$$

$$P(Y > 6) = 1 - P(Y \leq 6) = 1 - F(6^+) = 1 - F(7) = 1 - 0,95 = 0,05.$$

■

PRZYKŁAD I.2.2

Korzystając z wartości dystrybuanty zmiennej losowej dyskretnej Y podanych w tabeli I.2.4, podać jej rozkład.

Tabela I.2.4
Wartości dystrybuanty zmiennej losowej dyskretnej Y

y_i	-4	-3	0	1	9	10	$Y > 10$
$F(y_i)$	0,00	0,30	0,45	0,85	0,90	0,95	1,00

Rozwiązanie

Również i w tym przykładzie korzystamy ze wzoru (I.2.3), zakładając, że zmienna losowa Y przyjmuje wyłącznie wartości podane w tabeli I.2.4.

Oto obliczenia kilku wybranych prawdopodobieństw:

$$P(-4) = P(Y < -3) = F(-3) = 0,30,$$

$$P(-3) = P(Y < 0) - P(Y < -3) = F(0) - F(-3) = 0,45 - 0,30 = 0,15,$$

$$P(10) = P(Y \leq 10) - P(Y < 10) = F(Y > 10) - F(10) = 1 - 0,95 = 0,05.$$

Pełen rozkład zmiennej losowej Y przedstawiono w tabeli I.2.5.

Tabela I.2.5
Rozkład zmiennej losowej dyskretnej Y

y_i	-4	-3	0	1	9	10
p_i	0,30	0,15	0,40	0,05	0,05	0,05

■

I.2.2. Jednowymiarowa zmienna losowa ciągła

W zagadnieniach przyrodniczych i technicznych istnieje wiele zjawisk, do których opisu konieczne jest wprowadzenie zmiennej losowej, mogącej przyjmować dowolną wartość z określonego przedziału, a zatem mającej nieprzeliczalnie wiele wartości.

Definicja I.2.4. Zmienna losowa X , o dystrybuancie F jest **ciągła, jeżeli istnieje nieujemna funkcja f taka, że zachodzi następująca zależność:**

$$f \geq 0 \quad \forall_{x_0 \in R} \quad F(x_0) = \int_{-\infty}^{x_0} f(x) dx.$$

Funkcję f nazywamy **gęstością prawdopodobieństwa** zmiennej losowej X . Podobnie jak dla zmiennej skokowej, mamy wzajemnie jednoznaczną odpowiedniość pomiędzy rozkładem i dystrybucją. Zatem rozkład prawdopodobieństwa dla zmiennej ciągłej może być określony za pomocą funkcji dystrybuanty F lub funkcji gęstości f .

Własności zmiennej losowej ciągłej:

(1) f ciągła w $x \Rightarrow f(x) = F'(x)$, czyli w punktach ciągłości funkcji gęstości jej wartości równają się pochodnej dystrybuanty;

(2) $\int_{-\infty}^{+\infty} f(x) dx = 1$, czyli pole powierzchni pod wykresem funkcji gęstości w całym zbiorze liczb rzeczywistych wynosi 1;

(3) $P(X < x_0) = \int_{-\infty}^{x_0} f(x) dx$ z definicji I.2.2 oraz I.2.4.

PRZYKŁAD I.2.3

Zmienna losowa ciągła X jest określona w przedziale $(a, \sqrt{2})$. Funkcja gęstości dla niej ma postać $f(x) = 2x$. Wyznaczyć a – lewy koniec przedziału określoności zmiennej X , podać postać dystrybuanty F , narysować wykresy funkcji $f(x)$ i $F(x)$ oraz obliczyć następujące prawdopodobieństwa:

$$P(|X| < 1, 1), \quad P(1, 2 \leq X < 1, 5).$$

Rozwiązanie

Wiedząc, że zmienna X jest określona w przedziale $(a, \sqrt{2})$, z własności (2) zmiennej losowej ciągłej mamy:

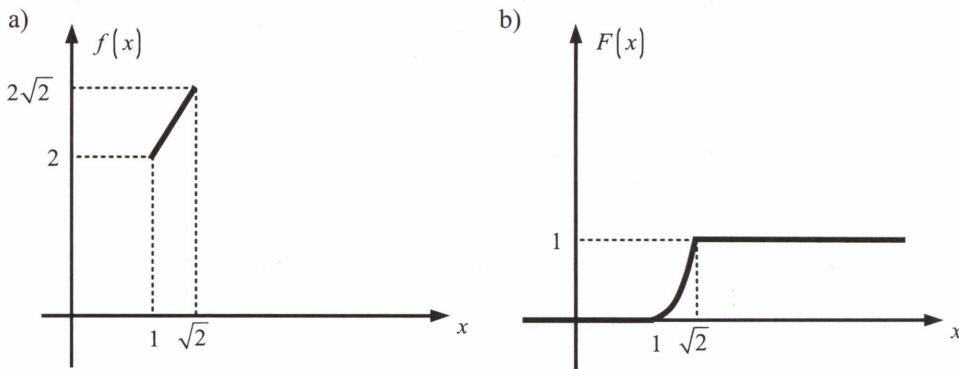
$$\int_{-\infty}^{+\infty} f(x) dx = \int_a^{\sqrt{2}} 2x dx = [x^2]_a^{\sqrt{2}} = 2 - a^2 = 1.$$

Stąd $a = 1$ lub $a = -1$. Jednak z definicji I.2.4 wiemy, że f jest funkcją nieujemną. Ponieważ $f(-1) = -2 < 0$, więc $a = 1$.

Zatem przedział wartości zmiennej X jest następujący $(1, \sqrt{2})$. Również z definicji I.2.4 wyznaczamy dystrybuantę $F(x)$:

$$F(x_0) = \int_{-\infty}^{x_0} f(x) dx = \int_1^{x_0} 2x dx = [x^2]_1^{x_0} = x_0^2 - 1, \quad \text{czyli:} \quad F(x) = x^2 - 1.$$

Na rysunku I.2.3 przedstawione są wykresy funkcji gęstości f oraz funkcji dystrybuanty F .



Rys. I.2.3. Wykres: a) funkcji gęstości f ; b) dystrybuanty F

Pozostają do obliczenia dwa prawdopodobieństwa. Korzystając z własności zmiennej losowej ciągłej oraz z wyznaczonego przedziału określoności dla X , otrzymujemy:

$$\begin{aligned} \text{a) } P(|X| < 1,1) &= P(X \in (-1,1; 1,1) \cap (1; \sqrt{2})) = P(X \in (1; 1,1)) = \\ &= \int_1^{1,1} f(x) dx = F(1,1) - F(1) = (1,1)^2 - 1 - (1^2 - 1) = 1,21 - 1 = 0,21; \end{aligned}$$

$$\begin{aligned} \text{b) } P(1,2 \leq X < 1,5) &= P(X \in [1,2; 1,5) \cap (1; \sqrt{2})) = P(X \in [1,2; \sqrt{2})) = \\ &= \int_{1,2}^{\sqrt{2}} f(x) dx = F(\sqrt{2}) - F(1,2) = \sqrt{2}^2 - 1 - ((1,2)^2 - 1) = 2 - 1,44 = 0,56. \end{aligned}$$

■

I.2.3. Charakterystyki liczbowe rozkładu zmiennej losowej

W celu scharakteryzowania zmiennej losowej (jej rozkładu) wyznacza się dla niej pewne parametry liczbowe, opisujące np. jej wartość przeciętną, rozrzut jej wartości, kształt wykresu prawdopodobieństwa itp. Najważniejszym parametrem jest **wartość przeciętna** zmiennej losowej, zwana też wartością średnią, wartością oczekiwaną lub nadzieją matematyczną.

Definicja I.2.5. Wartość przeciętna $E(X)$ zmiennej losowej X , to wartość następującego wyrażenia:

$$E(X) = \begin{cases} \sum_i x_i \cdot p_i & \text{dla zmiennej dyskretnej,} \\ \int_{-\infty}^{+\infty} x \cdot f(x) dx & \text{dla zmiennej ciągłej,} \end{cases}$$

jeśli powyższy szereg i całka są bezwzględnie zbieżne.

Korzystając z powyższej definicji, bardzo łatwo udowodnić następujące własności wartości przeciętnej dla $a = \text{constans}$:

- (1) $E(a) = a$
- (2) $E(a \cdot X) = a \cdot E(X)$
- (3) $E(X + a) = E(X) + a$
- (4) $E(X + Y) = E(X) + E(Y)$
- (5) $E(X \cdot Y) = E(X) \cdot E(Y)$, jeżeli X i Y są niezależne (pojęcie niezależności zmiennych losowych jest wyjaśnione w podrozdz. I.3.3).

(I.2.4)

Wartość przeciętna jest podstawą do wyliczenia wielu innych parametrów, m.in. tzw. momentów.

Definicja I.2.6. Momentem zwykłym rzędu r nazywamy liczbę $E[X^r]$. Natomiast moment centralny rzędu r to liczba $E[(X - E(X))^r]$.

Korzystając z definicji I.2.5 wartości przeciętnej, można napisać wzory pozwalające obliczyć momenty zwykłe i centralne (tab. I.2.6) dowolnego rzędu dla zmiennej dyskretnej i zmiennej ciągłej.

Tabela I.2.6
Momenty zwykłe i centralne zmiennej losowej

Momenty	Zmienna dyskretna	Zmienna ciągła
Moment zwykły rzędu r	$E[X^r] = \alpha_r = \sum_i x_i^r \cdot p_i$	$E[X^r] = \alpha_r = \int_{-\infty}^{+\infty} x^r \cdot f(x) dx$
Moment centralny rzędu r	$E[X - E(X)]^r = \mu_r = \sum_i (x_i - E(X))^r \cdot p_i$	$E[X - E(X)]^r = \mu_r = \int_{-\infty}^{+\infty} (x - E(X))^r \cdot f(x) dx$

ozn.
= informuje o przyjęciu dla definiowanego pojęcia, oznaczenia następującego po tym symbolu.

Zauważmy, że $\alpha_1 = E[X^1] = E(X)$, czyli moment zwykły rzędu pierwszego, to nie innego jak wartość przeciętna. Swoją odrębną nazwę ma również μ_2 – moment centralny rzędu drugiego, zwany **wariancją** i oznaczany $V(X)$. Wartość wariancji niesie informację na temat rozproszenia wartości zmiennej losowej. Pierwiastek z wariancji, zwany **odchyleniem standardowym**, oznaczamy $\sigma_x = \sigma(X) = \sqrt{V(X)}$. Stanowi on bardziej intuicyjną miarę rozproszenia, gdyż podawany jest w tych samych jednostkach co zmienna X .

W przypadku rozkładów symetrycznych wszystkie momenty centralne nieparzystych rzędów są równe zero. Dla zmiennych dyskretnych wynika to z faktu występowania wśród różnic $x_i - E(X)$ wyłącznie symetrycznych par wielkości różniących się jedynie znakiem, co przy nieparzystej potędze r : $(x_i - E(X))^r$ powoduje ich redukowanie się podczas sumowania. W przypadku zmiennych ciągłych wynika to z jednakowych wartości pól powierzchni pod wykresem funkcji podcałkowej $(x - E(X))^r \cdot f(x)$, znajdujących się nad i pod poziomą osią układu współrzędnych X .

Własności wariancji dla $a = \text{constans}$:

- (1) $V(a) = 0$
 - (2) $V(a \cdot X) = a^2 \cdot V(X)$
 - (3) $V(X + a) = V(X)$
 - (4) $V(X \pm Y) = V(X) + V(Y)$, jeżeli X i Y są niezależne
 - (5) $V(X) = E(X^2) - E^2(X)$
 - (6) $V(X) = E(X - a)^2 - (a - E(X))^2$
- (I.2.5)

Warto jeszcze zwrócić uwagę na trzy inne charakterystyki zmiennej losowej:

- 1) **współczynnik asymetrii** (skośność), badający asymetrię rozkładu $\gamma = \frac{\mu_3}{\sigma^3}$;
- 2) **współczynnik skupienia** (tzw. kurtოza), opisujący skoncentrowanie rozkładu $K = \frac{\mu_4}{\sigma^4}$;
- 3) **współczynnik spłaszczenia** (tzw. eksces) $E = K - 3$.

Trzy wymienione współczynniki są wielkościami niemianowanymi. Dzięki temu możliwe jest porównywanie rozkładów cech o różnych mianach ze względu na asymetrię i skupienie.

W analizach statystycznych wykorzystuje się jeszcze jedno pojęcie, odnoszące się bezpośrednio do wartości zmiennej. **Kwantyl** x_p rzędu p – jest to taka wartość zmiennej losowej (po uszeregowaniu jej w ciąg niemalejący), dla której spełnione jest:

$$F(x_p) = \sum_{x_i < x_p} p_i \leq p \leq \sum_{x_i \leq x_p} p_i = F(x_p^+) \quad (I.2.6a)$$

gdzie x_p^+ oznacza następną w kolejności wartość (liczbę) w zbiorze wartości zmiennej losowej.

Dla zmiennej ciągłej, ze względu na niemożność ustalenia elementu „następnego”, definicja kwantyla przyjmuje uproszczoną formę:

$$F(x_p) = p \quad (I.2.6b)$$

PRZYKŁAD I.2.4

Niech będzie dany rozkład zmiennej losowej skokowej X (dwa pierwsze wiersze w tabeli I.2.7).

Tabela I.2.7
Rozkład i wartości dystrybuanty zmiennej losowej skokowej X

x_i	-5	-1	1	2	7	$x > 7$
p_i	0,15	0,25	0,10	0,30	0,20	$\sum_i p_i = 1$
$F(x_i)$	0,00	0,15	0,40	0,50	0,80	1,00

W ostatnim wierszu zapisano wartości dystrybuanty, obliczone zgodnie z definicją I.2.2. Dla tego rozkładu, kwantyle wybranych rzędów p wynoszą:

$$p = 0,15, \quad x_p = -1, \text{ ponieważ } F(-1) = 0,15 \leq 0,15 \leq F(1) = 0,40;$$

następną po -1 wartością zmiennej losowej X w powyższym rozkładzie jest 1, więc:

$$p = 0,25, \quad x_p = -1, \text{ ponieważ } F(-1) = 0,15 \leq 0,25 \leq F(1) = 0,40,$$

$$p = 0,40, \quad x_p = 1, \text{ ponieważ } F(1) = 0,40 \leq 0,40 \leq F(2) = 0,50. \quad \blacksquare$$

PRZYKŁAD I.2.5

Niech rozkład zmiennej losowej ciągłej, określonej na przedziale $[0; 2]$, będzie zdefiniowany następującą funkcją gęstości: $f(x) = \frac{x}{2}$. Zgodnie z definicją I.2.4. dystrybuanta

rozkładu ma postać $F(x) = \frac{x^2}{4}$. Kwantyle wybranych rzędów wynoszą:

$$p = 0,15, \quad x_p = \sqrt{0,6}, \quad \text{ponieważ} \quad F(\sqrt{0,6}) = 0,15;$$

$$p = 0,25, \quad x_p = 1, \quad \text{ponieważ} \quad F(1) = 0,25;$$

$$p = 0,40, \quad x_p = \sqrt{1,6}, \quad \text{ponieważ} \quad F(\sqrt{1,6}) = 0,40.$$

Przykładowo moment zwykły czwartego rzędu tej zmiennej, na podstawie wzoru z tabeli I.2.6 wynosi:

$$\alpha_4 = E(X^4) = \int_{-\infty}^{+\infty} x^4 \cdot f(x) dx = \int_0^2 x^4 \cdot \frac{x}{2} dx = \int_0^2 \frac{x^5}{2} dx = \left[\frac{x^6}{12} \right]_0^2 = \frac{16}{3}.$$

■

PRZYKŁAD I.2.6

Wyprowadzić wzór na moment centralny rzędu trzeciego wyrażony za pomocą momentów zwykłych.

Rozwiązanie

Korzystamy z określenia momentu centralnego dowolnego rzędu (tab. I.2.6):

$$E[(X - E(X))^3] = \dots$$

Wykonujemy działanie podniesienia do potęgi wyrażenia w nawiasie:

$$\dots = E[X^3 - 3X^2 \cdot E(X) + 3X \cdot E^2(X) - E^3(X)] = \dots$$

Korzystamy z własności (I.2.4) liniowości wartości przeciętnej:

$$\dots = E[X^3] - 3E(X) \cdot E[X^2] + 3E^2(X) \cdot E(X) - E^3(X) \cdot E(1) = \dots$$

Wykonujemy działania algebraiczne na momentach zwykłych:

$$\dots = E[X^3] - 3E(X) \cdot E[X^2] + 3E^3(X) - E^3(X) = E[X^3] - 3E(X) \cdot E[X^2] + 2E^3(X) = \dots$$

Na koniec korzystamy z oznaczenia momentów zwykłych w tabeli I.2.6:

$$\dots = \alpha_3 - 3\alpha_1 \alpha_2 + 2\alpha_1^3.$$

■

PRZYKŁAD I.2.7

Korzystając z wartości dystrybuanty zmiennej losowej X podanych w tabeli I.2.8, policzyć moment centralny rzędu 2.

Tabela I.2.8Wartości dystrybuanty zmiennej losowej skokowej X

x_i	0	2	5	7	9	10
$F(x_i)$	0	0,25	0,60	0,75	0,90	0,95

Rozwiązanie

Korzystając z własności dystrybuanty (I.2.2), wyznaczamy rozkład zmiennej X , czyli prawdopodobieństwa p_i (tab. I.2.9).

Tabela I.2.9Rozkład zmiennej losowej X

x_i	0	2	5	7	9	10
p_i	0,25	0,35	0,15	0,15	0,05	0,05

Następnie, wiedząc, że moment centralny rzędu 2 to wariancja zmiennej losowej, w celu jej obliczenia korzystamy z własności wariancji (I.2.5): $V(X) = E(X^2) - E^2(X)$, czyli wyrażenia jej za pomocą momentów zwykłych:

$$E(X) = \sum_i x_i \cdot p_i = 0 \cdot 0,25 + 2 \cdot 0,35 + 5 \cdot 0,15 + 7 \cdot 0,15 + 9 \cdot 0,05 + 10 \cdot 0,05 = 3,45,$$

$$E(X^2) = \sum_i x_i^2 \cdot p_i = 0^2 \cdot 0,25 + 2^2 \cdot 0,35 + 5^2 \cdot 0,15 + 7^2 \cdot 0,15 + 9^2 \cdot 0,05 + 10^2 \cdot 0,05 = 21,55,$$

$$V(X) = 21,55 - (3,45)^2 = 9,6475. \quad \blacksquare$$

I.2.4. Wybrane rozkłady zmiennej losowej dyskretnej

W niniejszym podrozdziale zaprezentowane zostaną przykładowe rozkłady prawdopodobieństwa zmiennej losowej skokowej, wraz ze wzorami opisującymi ich podstawowe charakterystyki liczbowe.

Rozkład równomierny

W rozkładzie równomiernym prawdopodobieństwo przyjmowania przez zmienną losową dowolnej wartości jest jednakowe. Zatem dla n -elementowej przestrzeni zdarzeń elementarnych Ω rozkład wygląda jak w tabeli I.2.10.

Tabela I.2.10Rozkład równomierny zmiennej losowej skokowej X

x_i	x_1	x_2	...	x_n
p_i	$\frac{1}{n}$	$\frac{1}{n}$...	$\frac{1}{n}$

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i, \quad V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - E(X))^2 \quad (\text{I.2.7})$$

Rozkład jednopunktowy

Rozkład jednopunktowy jest szczególnym przypadkiem rozkładu równomiernego. Zmienna losowa przyjmuje tu tylko jedną wartość, a zatem jest to zdarzenie pewne (tab. I.2.11).

Tabela I.2.11
Rozkład jednopunktowy zmiennej losowej X

x_i	x
p_i	1

$$E(X) = x, \quad V(X) = 0 \quad (I.2.8)$$

Rozkład zero-jedynkowy z parametrem p (rozkład dwupunktowy)

Zmienna w rozkładzie dwupunktowym może przyjmować tylko dwie wartości. Opisujemy w ten sposób prawdopodobieństwo, że ustalone zdarzenie losowe zdarzy się lub nie (tab. I.2.12).

Tabela I.2.12
Rozkład dwupunktowy zmiennej losowej X

x_i	1	0
p_i	p	$q = 1 - p$

$$E(X) = p, \quad V(X) = p \cdot q, \quad \mu_3 = pq \cdot (1 - 2p) \quad (I.2.9)$$

Rozkład dwumianowy o parametrach n, p (rozkład Bernoullego)

Rozkład Bernoullego opisuje prawdopodobieństwo pojawienia się oczekiwanego wyniku (tzw. „sukcesu”) k razy w doświadczeniu powtarzanym niezależnie n razy, jeżeli znamy prawdopodobieństwo sukcesu p , w pojedynczej próbie. Funkcja prawdopodobieństwa zmiennej K o rozkładzie dwumianowym jest następującej postaci:

$$P(k; n, p) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, n, \quad \text{gdzie } q = 1 - p \quad (I.2.10)$$

$$E(K) = n \cdot p, \quad V(K) = n \cdot p \cdot q, \quad \mu_3 = npq \cdot (1 - 2p) \quad (I.2.11)$$

Ujemny rozkład dwumianowy o parametrach v, p (rozkład Pascala)

Zmienną losową K w rozkładzie Pascala jest liczba doświadczeń wykonanych według schematu Bernoullego, potrzebna do uzyskania v sukcesów. Prawdopodobieństwo uzyskania sukcesu w pojedynczym doświadczeniu wynosi p .

Funkcja prawdopodobieństwa zmiennej K jest postaci:

$$P(k; v, p) = \binom{k-1}{v-1} p^v q^{k-v}, \quad k = v, v+1, \dots, \quad \text{gdzie } q = 1-p \quad (\text{I.2.12})$$

$$E(K) = \frac{v}{p}, \quad V(K) = \frac{v \cdot q}{p^2}, \quad \mu_3 = \frac{vq \cdot (2-p)}{p^3} \quad (\text{I.2.13})$$

Rozkład geometryczny z parametrem p

Rozkład geometryczny jest szczególnym przypadkiem rozkładu Pascala dla $v = 1$ i zwany jest rozkładem czasu oczekiwania na pierwszy sukces. Funkcja prawdopodobieństwa jest postaci:

$$P(k; p) = p \cdot q^{k-1}, \quad k = 1, 2, 3, \dots, \quad \text{gdzie } q = 1-p \quad (\text{I.2.14})$$

$$E(K) = \frac{1}{p}, \quad V(K) = \frac{q}{p^2}, \quad \mu_3 = \frac{q \cdot (2-p)}{p^3} \quad (\text{I.2.15})$$

Rozkład Poissona z parametrem $\lambda > 0$

Funkcja prawdopodobieństwa zmiennej losowej K w rozkładzie Poissona jest postaci:

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (\text{I.2.16})$$

$$E(K) = \lambda, \quad V(K) = \lambda, \quad \mu_3 = \lambda \quad (\text{I.2.17})$$

Jest to rozkład graniczny dla ciągu rozkładów dwumianowych przy $n \rightarrow \infty$ oraz $n \cdot p_n \rightarrow \lambda$, czyli: $\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$, gdzie (n, p_n) – parametry zmiennej K o rozkładzie Bernoullego.

Dla $n \geq 50, p \leq 0,1, np \leq 10$ można wystarczająco dobrze przybliżyć rozkład dwumianowy rozkładem Poissona: $\binom{n}{k} p^k q^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda = np.$

PRZYKŁAD I.2.8

Niezawodność przyrządu (ustalona na podstawie badań laboratoryjnych) w pojedynnym pomiarze wynosi 0,9. Wykonano trzy pomiary tym przyrządem. Niech K będzie liczbą wyników pomiarów uznanych za bezbłędne, tzn. takich, których dokładność mieści się w dopuszczalnych granicach. Wyznaczyć rozkład zmiennej K , obliczyć moment zwykły rzędu 4 oraz sprawdzić symetrię rozkładu tej zmiennej. Obliczyć prawdopodobieństwo, że co najmniej dwa pomiary są błędne.

Rozwiązanie

Z treści zadania wnioskujemy, że chodzi o rozkład **Bernoullego**, gdyż mamy do czynienia z kilkukrotnym powtórzeniem tego samego doświadczenia (wykonanie pomiaru) oraz znamy prawdopodobieństwo sukcesu (bezbłędny pomiar) w pojedynczym doświadczeniu. Zatem, na podstawie oznaczeń przyjętych dla tego rozkładu, wypisujemy jego parametry

$$p = 0,9, \quad q = 1 - p = 0,1, \quad n = 3.$$

Prawdopodobieństwo uzyskania określonej liczby sukcesów k , we wszystkich n próbach, jest zdefiniowane następująco (zgodnie z (I.2.10)):

$$P(k; n, p) = \binom{n}{k} p^k q^{n-k}.$$

Zatem prawdopodobieństwo, że nie uzyskamy żadnego sukcesu, co w praktyce oznacza, że błędy wszystkich pomiarów będą przekraczały dopuszczalną wartość, wynosi:

$$P(0; 3; 0,9) = \binom{3}{0} 0,9^0 0,1^{3-0} = \frac{3!}{0!(3-0)!} \cdot 1 \cdot 0,1^3 = \frac{3!}{1 \cdot 3!} \cdot 0,001 = 0,001.$$

Podobnie wyznaczamy prawdopodobieństwa dla pozostałych $k = 1, 2, 3$, otrzymując rozkład zmiennej losowej K (tab. I.2.13):

Tabela I.2.13

Przykładowy rozkład Bernoullego zmiennej losowej K

k_i	0	1	2	3
p_i	0,001	0,027	0,243	0,729

Moment zwykły rzędu 4:

$$\alpha_4 = E(X^4) = \sum_i k_i^4 \cdot p_i = 0^4 \cdot 0,001 + 1^4 \cdot 0,027 + 2^4 \cdot 0,243 + 3^4 \cdot 0,729 = 62,964.$$

Z tabeli rozkładu I.2.13 widać wyraźnie, że rozkład nie jest symetryczny, ale chcąc to potwierdzić analitycznie oraz zbadać stopień asymetrii, wyznaczmy współczynnik asymetrii γ , zdefiniowany w podrozdziale I.2.3:

$$\gamma = \frac{\mu_3}{\sigma^3}.$$

Wykorzystując wzory (I.2.11) na moment centralny rzędu 3 i wariancję, otrzymujemy:

$$\gamma = \frac{npq(1-2p)}{\sqrt{npq}^3} = \frac{1-2p}{\sqrt{npq}} = \frac{1-2 \cdot 0,9}{\sqrt{3 \cdot 0,9 \cdot 0,1}} = \frac{-0,8}{\sqrt{0,27}} = -1,54.$$

Ujemna wartość γ świadczy o lewostronnej skośności rozkładu.

Pozostaje wyznaczyć interesujące nas prawdopodobieństwo.

Zdarzenie polegające na tym, że co najmniej dwa pomiary są błędne, oznacza w tym przypadku, że błędy dwóch lub trzech pomiarów nie mieszczą się w dopuszczalnych granicach. Ze względu na sposób zdefiniowania zmiennej losowej K jako liczby pomiarów bezbłędnych, nasze zdarzenie odpowiada sytuacji, w której mamy jeden lub zero pomiarów uznanych za bezbłędne.

Zatem zadanie sprowadza się do obliczenia prawdopodobieństwa:

$$P(K = 1 \vee K = 0) = 0,027 + 0,001 = 0,028,$$

czyli przeciętnie w 28 przypadkach na 1000, wykonując podobne doświadczenie trzech jednakowych pomiarów tym samym przyrządem, otrzymamy co najmniej dwa wyniki niemieszczące się w granicach dopuszczalnego błędu.

■

PRZYKŁAD I.2.9

Prawdopodobieństwo awarii urządzenia badawczego w jednym doświadczeniu wynosi $p = 0,01$. Doświadczenia można przeprowadzać dowolną liczbę razy. Przy założeniu, że są one niezależne, obliczyć prawdopodobieństwo, że druga awaria zdarzy się:

- a) w jedenastym doświadczeniu,
- b) najpóźniej w dziewiątym doświadczeniu,
- c) nie wcześniej niż w dziesiątym doświadczeniu.

Rozwiązanie

Mamy tu do czynienia z rozkładem **Pascala**. Zmienna losowa K opisuje liczbę doświadczeń, które trzeba wykonać, by uzyskać $v = 2$ „sukcesy” w postaci dwóch awarii, w określonym czasie. Zatem funkcja prawdopodobieństwa przyjmuje postać (zgodnie z (I.2.12)):

$$P(k; v=2, p=0,01) = \binom{k-1}{2-1} p^2 q^{k-2} = (k-1) \cdot 0,01^2 \cdot 0,99^{k-2}, \quad k = 2, 3, 4, \dots$$

Szukane prawdopodobieństwa wynoszą:

a) $P(K = 11) = 10 \cdot 0,0001 \cdot 0,99^9 \approx 0,00091$.

b) $P(K \leq 9) = P(2 \leq K \leq 9) = \sum_{k=2}^9 (k-1) \cdot 0,01^2 \cdot q^{k-2} = 0,0001 \cdot \sum_{k=2}^9 (k-1) \cdot q^{k-2} =$

$$= 10^{-4} \cdot \sum_{k=1}^8 k \cdot q^{k-1} = 10^{-4} \cdot \sum_{k=1}^8 \frac{\partial}{\partial q} q^k = 10^{-4} \cdot \frac{\partial}{\partial q} \sum_{k=1}^8 q^k = \dots$$

Korzystamy ze wzoru na sumę skończonego ciągu geometrycznego:

$$\begin{aligned} \dots &= 10^{-4} \cdot \frac{\partial}{\partial q} \left(q \cdot \frac{1-q^8}{1-q} \right) = 10^{-4} \cdot \left[\frac{1-q^8}{1-q} + q \cdot \frac{-8q^7(1-q) + 1-q^8}{(1-q)^2} \right] = \\ &= 10^{-4} \cdot \frac{1-q^8 - q + q^9 - 8q^8 + 8q^9 + q - q^9}{(1-q)^2} = 10^{-4} \cdot \frac{8q^9 - 9q^8 + 1}{(1-q)^2} = \\ &= 10^{-4} \cdot \frac{q^8 \cdot (8q - 9) + 1}{p^2} = 10^{-4} \cdot \frac{0,99^8 \cdot (8 \cdot 0,99 - 9) + 1}{10^{-4}} \approx 0,003436. \end{aligned}$$

$$c) P(K \geq 10) = 1 - P(K < 10) = 1 - P(K \leq 9) = 1 - 0,003436 = 0,996564.$$

■

PRZYKŁAD I.2.10

Firma ubezpieczeniowa ocenia, że co roku 1% ubezpieczonych traci życie w wypadkach określonego rodzaju. Jakie jest prawdopodobieństwo, że w danym roku firma będzie musiała wypłacić odszkodowanie więcej niż trzy razy, jeżeli od danego rodzaju wypadków ubezpieczyła 100 osób?

Rozwiązanie

Zmienną losową K w tym zagadnieniu jest liczba ubezpieczonych, którzy zginęli w danym roku. K ma rozkład **dwumianowy** z parametrami $p = 0,01$ i $n = 100$. Ze względu na wartości tych parametrów spełniające warunki: $n \geq 50$, $p \leq 0,1$, $np \leq 10$, można wystarczająco dobrze przybliżyć rozkład dwumianowy rozkładem **Poissona** (I.2.16):

$$\binom{n}{k} p^k q^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda = np.$$

Pytamy o następujące prawdopodobieństwo:

$$\begin{aligned} P(K > 3) &= 1 - P(K \leq 3) = 1 - P(0) - P(1) - P(2) - P(3) = \\ &= 1 - e^{-1} \left[\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!} + \frac{1^3}{3!} \right] = 1 - e^{-1} \left[1 + 1 + \frac{1}{2} + \frac{1}{6} \right] = 1 - \frac{1}{e} \cdot \frac{8}{3} = 1 - 0,981 = 0,019. \end{aligned}$$

Licząc to samo prawdopodobieństwo z rozkładu Bernoullego, otrzymujemy:

$$\begin{aligned} P(K > 3) &= 1 - P(K \leq 3) = 1 - P(0) - P(1) - P(2) - P(3) = 1 - \binom{100}{0} \cdot 0,01^0 \cdot 0,99^{100} - \\ &\quad - \binom{100}{1} \cdot 0,01^1 \cdot 0,99^{99} - \binom{100}{2} \cdot 0,01^2 \cdot 0,99^{98} - \binom{100}{3} \cdot 0,01^3 \cdot 0,99^{97} = \\ &= 1 - 1 \cdot 1 \cdot 0,99^{100} - 100 \cdot 0,01 \cdot 0,99^{99} - 50 \cdot 99 \cdot 0,0001 \cdot 0,99^{98} - 50 \cdot 33 \cdot 98 \cdot 0,000001 \cdot 0,99^{97} = \\ &= 1 - 0,366 - 0,370 - 0,185 - 0,061 = 0,018. \end{aligned}$$

Widać, że z wystarczającym przybliżeniem otrzymujemy podobny wynik, a rozkład Poissona zdecydowanie upraszcza obliczenia.

■

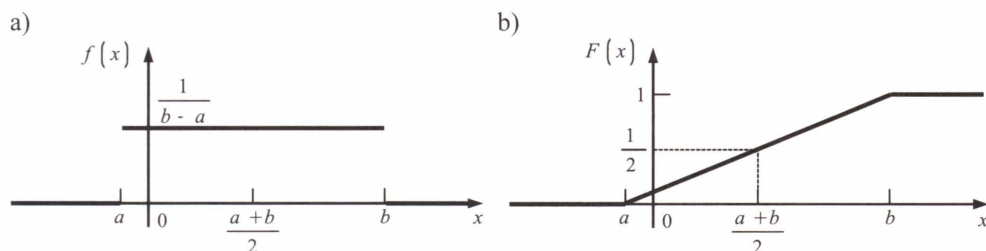
I.2.5. Wybrane rozkłady zmiennej losowej ciągłej

Po rozkładach zmiennych dyskretnych, zostaną zaprezentowane przykładowe rozkłady prawdopodobieństwa zmiennej losowej ciągłej, wraz z wzorami na ich podstawowe charakterystyki liczbowe.

Rozkład jednostajny (równomierny, prostokątny) na odcinku $[a; b]$

Na podstawie nazwy można wnioskować, że funkcja gęstości tego rozkładu będzie funkcją stałą w ustalonym przedziale $[a, b]$. Z uwagi na własności opisane w podrozdziale I.2.2., funkcja gęstości f oraz funkcja dystrybuanty F rozkładu prostokątnego (rys. I.2.4), skoncentrowanego w przedziale $[a; b]$, są następującej postaci:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla } x \in [a; b], \\ 0 & \text{dla } x \notin [a; b], \end{cases} \quad F(x) = \begin{cases} 0 & \text{dla } x \leq a \\ \frac{x-a}{b-a} & \text{dla } x \in [a; b] \\ 1 & \text{dla } x > b \end{cases} \quad (\text{I.2.18})$$



Rys. I.2.4. Gęstość rozkładu równomiernego skoncentrowanego na $[a; b]$ (a); dystrybuanta rozkładu równomiernego skoncentrowanego na $[a; b]$ (b)

Czytelnikowi pozostawiam, jako proste ćwiczenie, wyprowadzenie wzorów na podstawowe parametry tego rozkładu, takie jak wartość oczekiwana, wariancja itp.

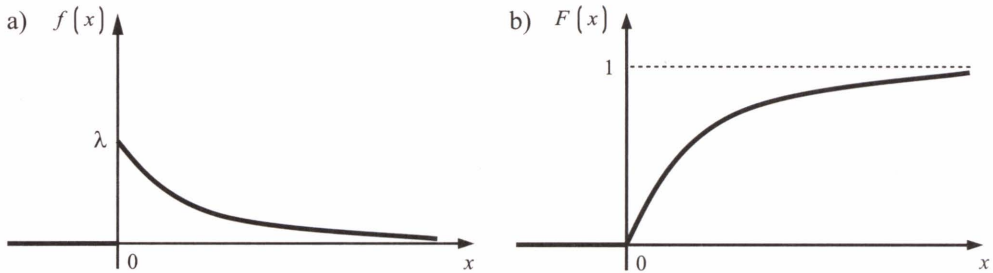
Prawdopodobieństwo, że zmienna losowa X o rozkładzie jednostajnym przyjmuje wartości w przedziale $[x_0; x_0 + h]$, przy założeniu, że $h > 0$ oraz x_0 i $x_0 + h$ należą do $[a; b]$, nie zależy od x_0 i jest wprost proporcjonalne do h . Własność ta umożliwia liczne zastosowania tego rozkładu (patrz przykład I.2.11).

Rozkład wykładniczy z parametrem $\lambda > 0$

Funkcja gęstości f oraz funkcja dystrybuanty F rozkładu wykładniczego (rys. I.2.5) są następującej postaci:

$$f(x) = \begin{cases} \lambda \cdot \exp(-\lambda x) & \text{dla } x \geq 0, \\ 0 & \text{dla } x < 0, \end{cases} \quad F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases} \quad (\text{I.2.19})$$

$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2} \quad (\text{I.2.20})$$



Rys. I.2.5. Gęstość rozkładu wykładniczego z parametrem λ (a);
dystrybuanta rozkładu wykładniczego (b)

Rozkład wykładniczy opisuje m.in. czas bezawaryjnej pracy urządzenia lub jego elementu oraz czas oczekiwania na obsługę (patrz przykład I.2.13), np. w sklepie. Posiada on własność zwaną **brakiem pamięci**. Oznacza to np., że pozostały czas pracy elementu nie zależy od dotychczasowego czasu jego pracy.

Wyjaśnia to następujące przekształcenie:

$$\begin{aligned} P(X \geq a+b | X \geq a) &= \frac{P(X \geq a+b \wedge X \geq a)}{P(X \geq a)} = \frac{P(X \geq a+b)}{P(X \geq a)} = \\ &= \frac{\exp(-\lambda(a+b))}{\exp(-\lambda a)} = \exp(-\lambda b) = P(X \geq b) \end{aligned} \quad (\text{I.2.21})$$

Rozkład normalny o parametrach μ i σ

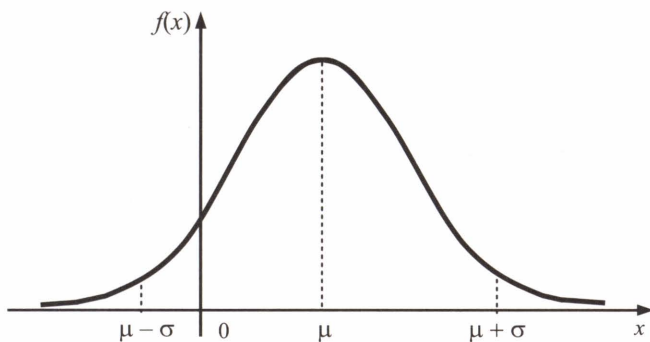
Fakt, że zmienna losowa X ma rozkład normalny o parametrach $\mu \in \mathbb{R}$ oraz $\sigma > 0$, będziemy symbolicznie zapisywać: $X \sim N(\mu, \sigma)$. Funkcja gęstości rozkładu normalnego ma postać:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{dla } x \in \mathbb{R} \quad (\text{I.2.22})$$

Ze wzoru (I.2.22) wynika, że $\forall_{x \in \mathbb{R}} f(x) > 0$. Zatem funkcja (I.2.22) spełnia warunek funkcji gęstości określony w definicji I.2.4. Dla zbadania przebiegu zmienności powyższej funkcji gęstości wyznaczamy pierwszą i drugą pochodną:

$$f'(x) = -\frac{x-\mu}{\sigma^3 \sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad f''(x) = -\frac{\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}{\sigma^3 \sqrt{2\pi}} \left[\frac{(x-\mu)^2}{\sigma^2} - 1\right] \quad (\text{I.2.23})$$

Miejsca zerowe pochodnych oraz ich znak określają ekstremum (maksimum) w punkcie $x = \mu$ oraz punkty przegięcia funkcji f w miejscach $x = \mu \pm \sigma$. Wykresem funkcji gęstości rozkładu normalnego Gaussa jest tzw. **krzywa Gaussa** (rys. I.2.6), dla której prosta $x = \mu$ jest osią symetrii. Wynika z tego, jak wyjaśniono w podrozdziale I.2.3, że wszystkie momenty centralne nieparzystego rzędu są równe 0. Dodatkowo, ponieważ funkcja gęstości rozkładu normalnego osiąga swoje maksimum w punkcie $x = \mu$, to wartość przeciętna tego rozkładu jest równa μ , gdyż jest ona interpretowana jako tzw. „wartość najbardziej prawdopodobna”, czyli o największym prawdopodobieństwie wystąpienia, a pole pod wykresem funkcji gęstości reprezentuje właśnie prawdopodobieństwo.



Rys. I.2.6. Gęstość rozkładu normalnego $N(\mu, \sigma)$

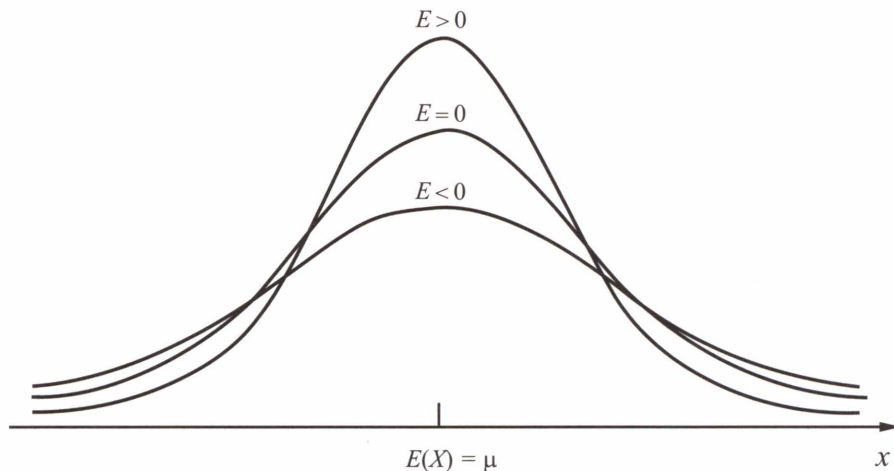
Można wyprowadzić ogólny wzór na momenty centralne parzystego rzędu. Przybiera on następującą postać:

$$E(X) = \mu, \quad \forall_{r \in \mathbb{N}^+} \mu_{2r-1} = 0 \quad \text{oraz} \quad \forall_{r \in \mathbb{N}^+} \mu_{2r} = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2r-1) \cdot \sigma^{2r} \quad (\text{I.2.24})$$

W szczególności wariancja: $V(X) = \mu_2 = \sigma^2$. Z powyższego wynikają również wartości współczynników asymetrii γ , skupienia K i spłaszczenia E :

$$\gamma = \frac{\mu_3}{\sigma^3} = 0, \quad K = \frac{\mu_4}{\sigma^4} = 3, \quad E = K - 3 = 0 \quad (\text{I.2.25})$$

Współczynnik spłaszczenia został tak zdefiniowany, by można było rozkład normalny traktować jako punkt odniesienia do badania skupienia i spłaszczenia dowolnego rozkładu. W większości wypadków, jeżeli $E > 0$, to rozkład jest bardziej skupiony (stromy) niż odpowiedni rozkład normalny (rys. I.2.7).



Rys. I.2.7. Gęstości rozkładów o różnych współczynnikach spłaszczenia E

Twierdzenie I.2.1. Jeżeli zmienną losową X o rozkładzie normalnym z parametrami μ i σ : $X \sim N(\mu, \sigma)$ poddamy standaryzacji, czyli przekształcimy następująco: $Z = \frac{X - \mu}{\sigma}$, to otrzymana zmienna Z ma tzw. **standaryzowany rozkład normalny** o parametrach 0 i 1: $Z \sim N(0; 1)$.

Dowód

Typ rozkładu dla zmiennej Z zostaje zachowany ze względu na liniowy charakter przekształcenia X w Z . Natomiast wartości parametrów dla zmiennej Z uzyskujemy dzięki przekształceniom wykorzystującym własności wartości przeciętnej (I.2.4) i wariancji (I.2.5).

Z własności (I.2.4) mamy:

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} \cdot E(X - \mu) = \frac{E(X) - E(\mu)}{\sigma} = \frac{\mu - \mu}{\sigma} = 0.$$

Podobnie z własności (I.2.5) mamy:

$$V(Z) = V\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \cdot V(X - \mu) = \frac{V(X) + V(\mu)}{\sigma^2} = \frac{\sigma^2 + 0}{\sigma^2} = 1.$$

W ten sposób uzyskujemy tezę twierdzenia $Z \sim N(0; 1)$.

c.b.d.u.

Dla zaznaczenia, że chodzi o standaryzowany rozkład normalny, funkcję gęstości będziemy oznaczać przez φ , a odpowiadającą jej dystrybuantę Φ :

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

$$\Phi(z_0) = P(z < z_0) = \int_{-\infty}^{z_0} \varphi(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} \exp\left(-\frac{z^2}{2}\right) dz \quad (1.2.26)$$

Uwaga. Warto zauważyć, że z symetrii funkcji $\varphi(z)$ względem pionowej osi układu współrzędnych OY , wynika równość $\Phi(-z) = 1 - \Phi(z)$. Wartości $\Phi(z)$ dla $z > 0$ są zawarte w tablicy 1 dystrybuanty rozkładu normalnego, zamieszczonej w dodatku na końcu książki.

Rozkład chi-kwadrat χ^2 o k stopniach swobody

Niech $Z_i \sim N(0, 1)$, wówczas suma kwadratów k niezależnych zmiennych losowych Z_i ma rozkład χ^2 o k stopniach swobody, czyli $\chi_k^2 = \sum_{i=1}^k Z_i^2$. Funkcja gęstości rozkładu χ^2 jest postaci:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}} \cdot \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right) & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases} \quad (1.2.27)$$

gdzie Γ jest to tzw. **funkcja specjalna gamma**, rozszerzająca pojęcie silni na zbiór liczb zespolonych.

Funkcja gamma wyraża się następującą całką:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} \cdot e^{-x} dx, \quad (1.2.28)$$

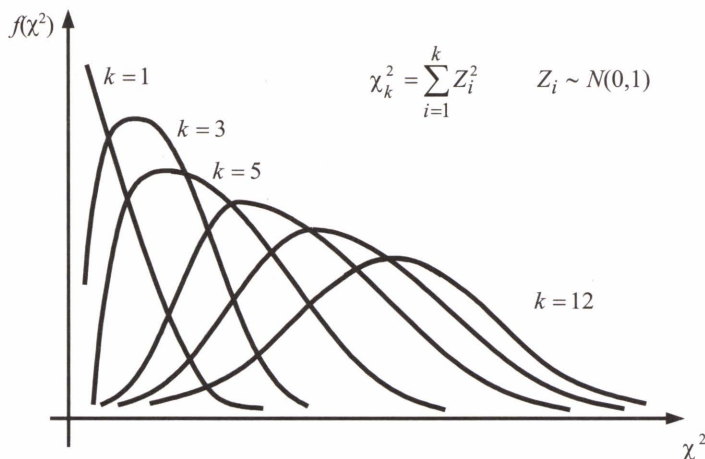
gdzie z – liczba zespolona o dodatniej części rzeczywistej $\text{Re}(z) > 0$.

Podstawowe charakterystyki liczbowe tego rozkładu są następujące:

$$E(X) = k, \quad V(X) = 2k, \quad \gamma = \frac{4}{\sqrt{2k}}, \quad E = \frac{12}{k} \quad (1.2.29)$$

Zauważmy, że $\lim_{k \rightarrow \infty} \gamma = 0$ oraz $\lim_{k \rightarrow \infty} E = 0$. Wynika stąd, że dla rosnącej liczebności pomiarów zmiennej losowej X , co pociąga za sobą wzrost liczby stopni swobody (tzw. obserwacji nadliczbowych), asymetria (zawsze prawostronna, bo: $\forall_{k \in N} \gamma > 0$) jest niwelowana. Rozkład zbliża się do symetrycznego.

Podobnie dzieje się ze współczynnikiem spłaszczenia. Z obu powyższych faktów można wyciągnąć wniosek, że przy znacznym wzroście k – rozkład χ^2 dąży do rozkładu normalnego, który jest symetryczny i dla którego eksces E wynosi 0 (rys. I.2.8).



Rys. I.2.8. Gęstość rozkładu χ^2 dla różnej liczby stopni swobody k

Chcąc uzyskać szybszą zbieżność z rozkładem normalnym, wystarczy zmienną χ_k^2 poddać następującemu przekształceniu $U^2 \stackrel{\text{ozn.}}{=} \sqrt{2 \cdot \chi_k^2}$. Wówczas $\lim_{k \rightarrow \infty} U^2 \sim N(\sqrt{2k-1}; 1)$.

Dowód pomijamy.

Na podstawie stosownego twierdzenia (Lindeberga–Levy’ego – twierdzenie II.3.2, w podrozd. II.3.1), dla $k > 50$ rozkład χ^2 można z dużą dokładnością aproksymować rozkładem normalnym.

Rozkład T-Studenta o k stopniach swobody

Niech $Z \sim N(0; 1)$; wtedy zmienna losowa zdefiniowana następująco:

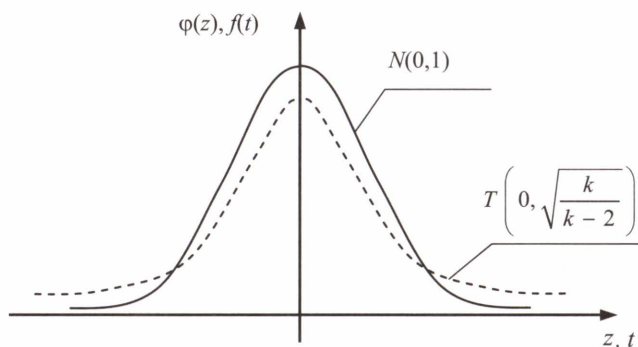
$$T_k = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}} \quad (\text{I.2.30})$$

ma **rozkład T-Studenta** o k stopniach swobody.

Funkcja gęstości tego rozkładu jest postaci:

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi} \cdot \Gamma\left(\frac{k}{2}\right) \left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \quad \text{dla } t \in R \quad (\text{I.2.31})$$

Z postaci funkcji gęstości widać, że jej wykres jest symetryczny względem prostej $t = 0$. Wynika to z faktu, że zmienna t występuje w funkcji (I.2.31) tylko w parzystej potęgze t^2 , które to wyrażenie przyjmuje jednakowe wartości dla równych co do wartości bezwzględnej ujemnych i dodatnich t . Zatem wartość przecięta wynosi 0. Wykres ten jest również podobny w swym kształcie do krzywej Gaussa. Na rysunku I.2.9 umieszczono krzywą Gaussa dla standaryzowanego rozkładu normalnego wraz z krzywą będącą obrazem funkcji (I.2.31). Dla dużych k ($k > 30$) rozkład Studenta można aproksymować rozkładem normalnym.



Rys. I.2.9. Rozkład Studenta (- -) w zestawieniu ze standaryzowanym rozkładem normalnym (—)

Z symetrii rozkładu Studenta wynika, że wszystkie momenty nieparzystych rzędów są równe 0. Natomiast wariancja zmiennej losowej T_k wynosi:

$$V(T) = \frac{k}{k-2} \quad \text{dla } k > 2 \quad (\text{I.2.32})$$

Rozkład F-Snedecora o k_1, k_2 stopniach swobody

Rozkładem **F-Snedecora** o k_1, k_2 stopniach swobody nazywamy rozkład ilorazu

$$F = \frac{X/k_1}{Y/k_2} \quad (\text{I.2.33})$$

gdzie X, Y to niezależne zmienne losowe, mające rozkłady chi-kwadrat odpowiednio o k_1 i k_2 stopniach swobody $X \sim \chi_{k_1}^2, Y \sim \chi_{k_2}^2$.

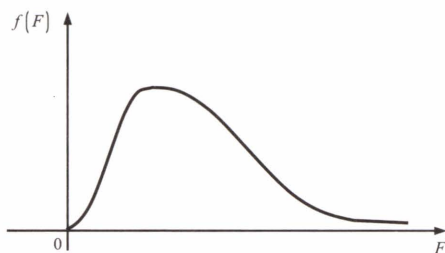
Funkcja gęstości tego rozkładu jest postaci:

$$f(t) = \begin{cases} \frac{\Gamma\left(\frac{k_1 + k_2}{2}\right) \left(\frac{r_1}{r_2}\right)^{\frac{k_1}{2}} \cdot t^{\frac{k_1-2}{2}}}{\Gamma\left(\frac{k_1}{2}\right) \Gamma\left(\frac{k_2}{2}\right) \left(1 + \frac{k_1}{k_2} t\right)^{\frac{k_1+k_2}{2}}} & \text{dla } t > 0 \\ 0 & \text{dla } t \leq 0 \end{cases} \quad (\text{I.2.34})$$

a wartość przeciętna i wariancja:

$$E(t) = \frac{k_2}{k_2 - 2}, \quad V(t) = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)} \quad (\text{I.2.35})$$

Na rysunku I.2.10 przedstawiono rozkład F-Snedecora o stopniach swobody $k_1 = 10$ i $k_2 = 10$.



Rys. I.2.10. Gęstość rozkładu $F_{10,10}$ dla stopni swobody $k_1 = 10$, $k_2 = 10$

PRZYKŁAD I.2.11

Na stację metra pociągi przyjeżdżają dokładnie co 4 minuty. Niech zmienna X opisuje czas oczekiwania na przybycie pociągu przez pasażera przybywającego na stację w przypadkowej chwili. Określić rozkład zmiennej X , jej dystrybuantę oraz obliczyć prawdopodobieństwo, że czas oczekiwania wynosi co najwyżej minutę.

Rozwiązanie

Zmienna losowa X ma rozkład **jednostajny**, zdefiniowany następującą funkcją gęstości:

$$f(x) = \begin{cases} \frac{1}{4} & \text{dla } x \in [0; 4], \\ 0 & \text{dla } x \notin [0; 4], \end{cases} \quad \text{a dystrybuanta ma postać} \quad F(x) = \begin{cases} 0 & \text{dla } x \leq 0, \\ \frac{x}{4} & \text{dla } x \in (0; 4], \\ 1 & \text{dla } x > 4. \end{cases}$$

Szukane prawdopodobieństwo wynosi:

$$P(X \leq 1) = \int_0^1 f(x) dx = \int_0^1 \frac{1}{4} dx = \frac{1}{4} = 0,25.$$

■

PRZYKŁAD I.2.12

Wiedząc, że zmienna X ma rozkład **normalny** o parametrach $\mu = 3,5$ oraz $\sigma = 4$, czyli $X \sim N(3,5; 4)$, zmienna T ma rozkład **Studenta** o $k = 17$ stopniach swobody, czyli $T \sim T_{17}$, a zmienna Y ma rozkład **chi-kwadrat** o $k = 7$ stopniach swobody, czyli $Y \sim \chi^2_7$ – zaznaczyć na rysunkach i obliczyć następujące prawdopodobieństwa:

- a) $P(-7,5 < X \leq 13,62)$, $P(-0,5 \leq X \leq 7,5)$;
- b) $P(|T| \geq 1,333)$;
- c) $P(|Y| < 1,69)$.

Rozwiązanie

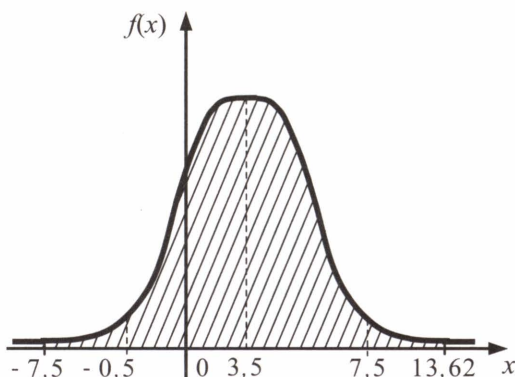
Teoretycznie można by obliczyć powyższe prawdopodobieństwa, całkując funkcje gęstości trzech rozpatrywanych rozkładów, tj. funkcje postaci (I.2.22), (I.2.31) i (I.2.27), podobnie jak przebiegało to w przykładzie I.2.11. Jednakże byłoby to zbyt pracochłonne. Dla każdego z tych rozkładów istnieją tablice statystyczne, z których zdecydowanie prościej będzie tu skorzystać.

Ad a)

Z całej rodziny rozkładów normalnych, w tablicach ujęto jedynie rozkład normalny standaryzowany. Zatem chcąc skorzystać z tablic, należy najpierw przeprowadzić standaryzację podanego rozkładu normalnego $N(3,5; 4)$, zgodnie z twierdzeniem I.2.1. W praktyce oznacza to, że należy znaleźć taki przedział $(z_1, z_2]$, dla zmiennej standaryzowanej Z , który będzie odpowiednikiem przedziału określonego dla zmiennej X w następującym sensie:

$$P(z_1 < Z \leq z_2) = P(-7,5 < X \leq 13,62).$$

Prawdopodobieństwo, które mamy obliczyć jako pierwsze, odpowiada polu powierzchni pod wykresem funkcji gęstości, zaznaczonemu na rysunku I.2.11.

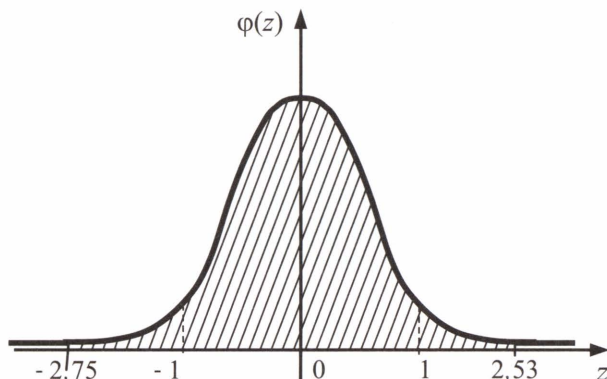


Rys. I.2.11. Graficzna reprezentacja prawdopodobieństwa tego, że zmienna losowa X o rozkładzie $N(3,5; 4)$ przyjmuje wartości w przedziale $(-7,5; 13,62]$

Z twierdzenia I.2.1 wiemy, że zależność pomiędzy odpowiadającymi sobie końcami obu przedziałów jest następująca:

$$z_1 = \frac{-7,5 - \mu}{\sigma} = \frac{-7,5 - 3,5}{4} = -2,75, \quad z_2 = \frac{13,62 - \mu}{\sigma} = \frac{13,62 - 3,5}{4} = 2,53.$$

Po standaryzacji poszukiwane pole powierzchni wygląda jak na rysunku I.2.12.



Rys. I.2.12. Graficzna reprezentacja prawdopodobieństwa tego, że standaryzowana zmienna losowa Z przyjmuje wartości w przedziale $(-2,75; 2,53]$

Podobnie jak w przykładzie I.2.3 prawdopodobieństwo możemy wyrazić za pomocą dystrybucyj:

$$P(-7,5 < X \leq 13,62) = P(-2,75 < Z \leq 2,53) = \Phi(2,53) - \Phi(-2,75) = \dots$$

Zgodnie z uwagą na stronie 35, ze względu na symetrię rozkładu normalnego standaryzowanego względem pionowej osi układu współrzędnych, dystrybucyjną dla ujemnej wartości zmiennej możemy zastąpić dystrybucyjną dla wartości dodatniej:

$$\dots = \Phi(2,53) - (1 - \Phi(2,75)) = \Phi(2,53) - 1 + \Phi(2,75) = \dots$$

Wartości dystrybucyjny odczytujemy z tablic dystrybucyjny standaryzowanego rozkładu normalnego, stąd ostatecznie prawdopodobieństwo wynosi:

$$\dots = 0,994297 - 1 + 0,997020 = 0,991317.$$

Analogicznie obliczymy drugie z zadanych w rozkładzie normalnym prawdopodobieństw:

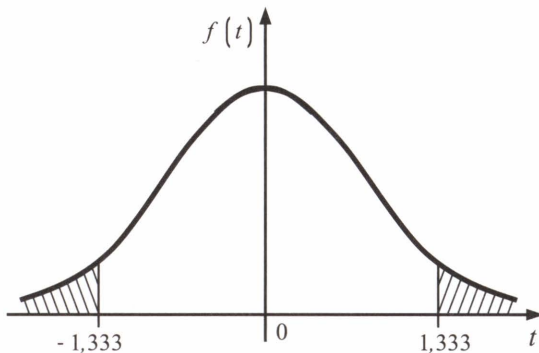
$$\begin{aligned} P(-0,5 \leq X \leq 7,5) &= P\left(\frac{-0,5 - 3,5}{4} \leq Z \leq \frac{7,5 - 3,5}{4}\right) = P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = \\ &= \Phi(1) - (1 - \Phi(1)) = \Phi(1) - 1 + \Phi(1) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0,841345 - 1 = 0,68269. \end{aligned}$$

Zatem prawdopodobieństwo $P(\mu - \sigma \leq X \leq \mu + \sigma)$, dla zadanego rozkładu normalnego, wynosi w przybliżeniu 0,68. Możemy udowodnić to dla dowolnych wartości μ oraz σ :

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} \leq Z \leq \frac{\mu + \sigma - \mu}{\sigma}\right) = P(-1 \leq Z \leq 1) \cong 0,68.$$

Ad b)

Poszukiwane prawdopodobieństwo w rozkładzie **Studenta** odpowiada zaznaczonemu na rysunku I.2.13 polu powierzchni pod wykresem funkcji gęstości.



Rys. I.2.13. Graficzna reprezentacja prawdopodobieństwa tego, że zmienna losowa T o rozkładzie T_{17} spełnia warunek $|T| \geq 1,333$

Ponieważ mamy do dyspozycji tablice kwantyli rozkładu **Studenta**, również w tym wypadku prawdopodobieństwo wyrazimy za pomocą dystrybucyj, która jest związana z kwantylem za pomocą określenia (I.2.6b).

Z tablic kwantylów rozkładu **Studenta** odczytujemy, że dla 17 stopni swobody 1,333 jest kwantylem rzędu 0,90: $t(0,90; 17) = 1,333$, czyli $F(1,333) = 0,90$ w tym rozkładzie.

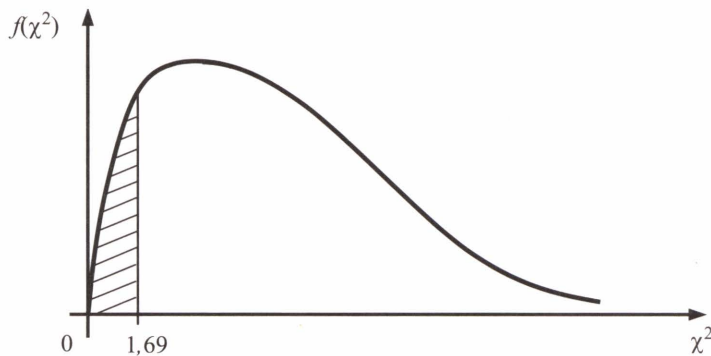
Zatem:

$$\begin{aligned} P(|T| \geq 1,333) &= 2 \cdot P(T \geq 1,333) = 2 \cdot (1 - P(T < 1,333)) = \\ &= 2 \cdot (1 - F(1,333)) = 2 \cdot (1 - 0,90) = 0,20. \end{aligned}$$

Ad c)

Poszukiwane prawdopodobieństwo w rozkładzie **chi-kwadrat** odpowiada zakreskowanemu polu powierzchni na rysunku I.2.14.

Podobnie jak w przykładzie b), chcemy skorzystać z tablic kwantylów, więc wyrażamy prawdopodobieństwo za pomocą dystrybucyj, pamiętając, że w rozkładzie **chi-kwadrat** zmienna losowa przyjmuje tylko wartości nieujemne. Z tablic kwantylów tego rozkładu dla siedmiu stopni swobody odczytujemy: $\chi^2(0,025; 7) = 1,69$, co oznacza kwantyl rozkładu **chi-kwadrat** dla prawdopodobieństwa $p = 0,025$ i $k = 7$ stopni swobody, czyli $F(1,69) = 0,025$.



Rys. I.2.14. Graficzna reprezentacja prawdopodobieństwa tego, że zmienna losowa Y o rozkładzie χ^2_7 spełnia warunek $|Y| < 1,69$

Zatem:

$$P(|Y| < 1,69) = P(Y \in [0; 1,69)) = F(1,69) = 0,025. \quad \blacksquare$$

PRZYKŁAD I.2.13

Czas między kolejnymi zgłoszeniami w centrali telefonicznej, mierzony w minutach, ma rozkład **wykładniczy** o parametrze $\lambda = 0,5$. Obliczyć średni czas między kolejnymi zgłoszeniami oraz prawdopodobieństwo, że zgłoszenie nastąpi do 4 minut.

Rozwiązanie

Ze wzoru (I.2.20) wynika, że wartość średnia zmiennej losowej, czyli średni czas między kolejnymi zgłoszeniami wynosi:

$$E(X) = \frac{1}{\lambda} = 2$$

Znając postać funkcji gęstości oraz dystrybucyjny (I.2.19) tego rozkładu wyznaczamy prawdopodobieństwo:

$$P(X \leq 4) = F(4) = 1 - \exp(-0,5 \cdot 4) = 1 - e^{-2} \approx 0,86. \quad \blacksquare$$

I.3. Zmienna losowa dwuwymiarowa

Ze zmienną losową dwuwymiarową, i ogólnie: wielowymiarową, mamy do czynienia, badając ustaloną populację ze względu na więcej niż jedną cechę. Na przykład analiza warunków atmosferycznych pod względem temperatury i wilgotności powietrza oraz ciśnienia atmosferycznego, czy też badanie jakości gleby ze względu na jej żyzność, urodzajność, odczyn, porowatość, wilgotność, zwięzłość itd.

Zmienną losową dwuwymiarową definiujemy następująco:

Definicja I.3.1. Niech X i Y – zmienne losowe, określone na tych samych lub różnych przestrzeniach probabilistycznych. Parę (X, Y) nazywamy **dwuwymiarową zmienną losową** lub **dwuwymiarowym wektorem losowym**.

Definicja I.3.2. Funkcja $F: R^2 \rightarrow [0, 1]$ jest **dystrybuantą dwuwymiarowej zmiennej losowej** (X, Y) jeżeli: $\forall_{(x_0, y_0) \in R^2} F(x_0, y_0) = P(X < x_0 \wedge Y < y_0)$.

Analogicznie definiujemy wielowymiarową zmienną losową i jej dystrybuantę.

Własności dystrybuanty dwuwymiarowej zmiennej losowej:

- (1) $\forall_{(x, y) \in R^2} F(x, y) \in [0, 1]$
- (2) $\forall_{x \in R} \lim_{y \rightarrow -\infty} F(x, y) = 0$ oraz $\forall_{y \in R} \lim_{x \rightarrow -\infty} F(x, y) = 0$
- (3) $\lim_{x \rightarrow +\infty} F(x, y) = 1$ (I.3.1)
- (4) $\forall_{\substack{x_1 \leq x_2 \\ y_1 \leq y_2}} P(X \in [x_1, x_2) \wedge Y \in [y_1, y_2)) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$
- (5) F jest funkcją niemalejącą i lewostronnie ciągłą ze względu na każdy z argumentów x lub y .

Podobnie jak dla zmiennej losowej jednowymiarowej, znając rozkład prawdopodobieństwa dwuwymiarowej zmiennej losowej, można jednoznacznie wyznaczyć jej dystrybuantę i odwrotnie.

I.3.1. Dwuwymiarowa zmienna losowa dyskretna

Definicja I.3.3. Dwuwymiarową zmienną losową (X, Y) , o co najwyżej przeliczalnym zbiorze wartości (x_i, y_j) , które są przyjmowane z prawdopodobieństwami:

$$P(X = x_i, Y = y_j) = p_{ij} \quad \text{dla} \quad i, j \in N$$

takimi, że $\sum_i \sum_j p_{ij} = 1$, nazywamy **dwuwymiarową zmienną losową skokową (dyskretną)**.

Rozkład prawdopodobieństwa dwuwymiarowej zmiennej losowej dyskretnej wygodnie jest zapisywać w formie tabeli (tab. I.3.1).

Tabela I.3.1

Rozkład dwuwymiarowej zmiennej losowej (X, Y)

$X \backslash Y$	y_1	y_2	...	y_m
x_1	p_{11}	p_{12}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2m}
x_3	p_{31}	p_{32}	...	p_{3m}
...
x_n	p_{n1}	p_{n2}	...	p_{nm}

PRZYKŁAD I.3.1

Korzystając z rozkładu dwuwymiarowej zmiennej losowej dyskretnej (X, Y) podanej w tabeli I.3.2, wyznaczyć wartości dystrybuanty, obliczyć i wyrazić za pomocą dystrybuanty następujące prawdopodobieństwa:

- $P(1 < X \leq 4 \wedge 2 \leq Y < 7)$,
- $P(2 \leq X < 4 \wedge Y > 2)$.

Tabela I.3.2

Przykładowy rozkład zmiennej losowej dwuwymiarowej (X, Y)

$X \backslash Y$	1	2	4	10
2	0,05	0,05	0,15	0,10
5	0,10	0,15	0,00	0,20
7	0,00	0,05	0,05	0,00
9	0,05	0,00	0,00	0,05

Rozwiązanie

Wartości dystrybuanty obliczamy na podstawie definicji I.3.2:

$$F(X = 1 \wedge Y = 2) = F(1, 2) = P(X < 1 \wedge Y < 2) = 0,$$

$$F(4, 2) = P(X < 4 \wedge Y < 2) = 0,$$

$$F(2, 7) = P(X < 2 \wedge Y < 7) = P(X = 1 \wedge (Y = 2 \vee Y = 5)) = 0,05 + 0,10 = 0,15.$$

Pełny wynik obliczeń zawiera tabela I.3.3.

Tabela I.3.3

Wartości dystrybuanty zmiennej losowej dwuwymiarowej (X, Y)

$\begin{matrix} X \\ Y \end{matrix}$	$(-\infty; 1]$	$(1; 2]$	$(2; 4]$	$(4; 10]$	$(10; +\infty)$
$(-\infty; 2]$	0,00	0,00	0,00	0,00	0,00
$(2; 5]$	0,00	0,05	0,10	0,25	0,35
$(5; 7]$	0,00	0,15	0,35	0,50	0,80
$(7; 9]$	0,00	0,15	0,40	0,60	0,90
$(9; +\infty)$	0,00	0,20	0,45	0,65	1,00

Prawdopodobieństwa wyznaczmy na podstawie rozkładu, a do zapisania ich za pomocą dystrybuanty wykorzystamy własność (4) dystrybuanty dwuwymiarowej zmiennej losowej (I.3.1).

- a) $P(1 < X \leq 4 \wedge 2 \leq Y < 7) = P((X = 2 \vee X = 4) \wedge (Y = 2 \vee Y = 5)) =$
 $= 0,05 + 0,15 + 0,15 + 0,00 = 0,35;$
 $P(1 < X \leq 4 \wedge 2 \leq Y < 7) = F(4^+, 7) - F(1^+, 7) - F(4^+, 2) + F(1^+, 2) =$
 $= F(10, 7) - F(2, 7) - F(10, 2) + F(2, 2) = 0,50 - 0,15 - 0,00 + 0,00 = 0,35;$
- b) $P(2 \leq X < 4 \wedge Y > 2) = P(X = 2 \wedge (Y = 5 \vee Y = 7 \vee Y = 9)) = 0,15 + 0,05 + 0,00 = 0,20;$
 $P(2 \leq X < 4 \wedge Y > 2) = F(4, 9^+) - F(2, 9^+) - F(4, 2^+) + F(2, 2^+) =$
 $= 0,45 - 0,20 - 0,10 + 0,05 = 0,20.$ ■

I.3.1.1. Rozkłady brzegowe zmiennych losowych dyskretnych

Ostatnia kolumna i ostatni wiersz tabeli I.3.4 zawierają tzw. rozkłady brzegowe, odpowiednio zmiennej X i zmiennej Y .

Rozkłady brzegowe określa się następująco:

Definicja I.3.4. Rozkład brzegowy zmiennej X w rozkładzie dwuwymiarowej zmiennej losowej dyskretnej (X, Y) , to wartości zmiennej X wraz z prawdopodobieństwami, przyporządkowanymi im w następujący sposób:

$$\begin{aligned} \text{ozn.} \\ p_{i \cdot} &= P(X = x_i) = P(X = x_i \wedge Y = y_1) + P(X = x_i \wedge Y = y_2) + \dots = \\ &= \sum_j P(X = x_i \wedge Y = y_j) = \sum_j p_{ij}. \end{aligned}$$

Analogicznie definiujemy rozkład brzegowy zmiennej Y :

$$\begin{aligned} \text{ozn.} \\ p_{\cdot j} &= P(Y = y_j) = P(X = x_1 \wedge Y = y_j) + P(X = x_2 \wedge Y = y_j) + \dots = \\ &= \sum_i P(X = x_i \wedge Y = y_j) = \sum_i p_{ij}. \end{aligned}$$

Tabela I.3.4

Rozkład dwuwymiarowej zmiennej losowej (X, Y) wraz z rozkładami brzegowymi zmiennych X i Y

$X \backslash Y$	y_1	y_2	...	y_m	$p_{x.}$
x_1	p_{11}	p_{12}	...	p_{1m}	$p_{1.} = \sum_{j=1}^m p_{1j}$
x_2	p_{21}	p_{22}	...	p_{2m}	$p_{2.} = \sum_{j=1}^m p_{2j}$
x_3	p_{31}	p_{32}	...	p_{3m}	$p_{3.} = \sum_{j=1}^m p_{3j}$
...
x_n	p_{n1}	p_{n2}	...	p_{nm}	$p_{n.} = \sum_{j=1}^m p_{nj}$
$p_{.y}$	$p_{.1} = \sum_{i=1}^n p_{i1}$	$p_{.2} = \sum_{i=1}^n p_{i2}$...	$p_{.m} = \sum_{i=1}^n p_{im}$	$\sum_{i,j} p_{ij} = 1$

Oczywiście spełnione są warunki rozkładu: $\sum_i p_{i.} = 1$ oraz $\sum_j p_{.j} = 1$.

Dystrybuanty rozkładów brzegowych zmiennej skokowej określamy następująco:

$$F_x(x_0) = \sum_{x < x_0} p_{i.}, \quad F_y(y_0) = \sum_{y < y_0} p_{.j} \quad (\text{I.3.2})$$

I.3.1.2. Rozkłady warunkowe zmiennych losowych dyskretnych

W podrozdziale I.1.2 została podana definicja prawdopodobieństwa warunkowego. Analogicznie definiuje się tzw. rozkłady warunkowe zmiennych losowych.

Definicja I.3.5. Rozkład warunkowy zmiennej losowej X , pod warunkiem, że zmienna Y przyjmuje ustaloną wartość y_j , określamy następująco:

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{.j}}.$$

Rozkład warunkowy zmiennej losowej Y , pod warunkiem, że zmienna X przyjmuje ustaloną wartość x_i , określamy następująco:

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i.}}.$$

Czytelnikowi pozostawiam sprawdzenie, że rozkład warunkowy rzeczywiście spełnia warunki rozkładu, zawarte w definicji I.2.3.

PRZYKŁAD I.3.2

Dla zmiennej losowej o rozkładzie z przykładu I.3.1 wyznaczyć rozkłady brzegowe zmiennych X i Y , dystrybuanty rozkładów brzegowych oraz obliczyć następujące prawdopodobieństwa warunkowe:

- $P(X=10|Y=7)$,
- $P((Y=5 \vee Y=2)|X=4)$.

Rozwiązanie

Rozkłady brzegowe wyznaczamy na podstawie definicji I.3.4, a dystrybuanty rozkładów brzegowych ze wzoru (I.3.2) i zamieszczamy jako dodatkowe wiersze (rozkład i dystrybuanta zmiennej X) i dodatkowe kolumny (rozkład i dystrybuanta zmiennej Y) w tabeli rozkładu dwuwymiarowej zmiennej losowej (X, Y) (tab. I.3.5).

Tabela I.3.5

Rozkład zmiennej losowej (X, Y) wraz z rozkładami brzegowymi zmiennych X, Y oraz wartościami dystrybuant w rozkładach brzegowych

$\begin{matrix} X \\ Y \end{matrix}$	1	2	4	10	p_y	F_y
2	0,05	0,05	0,15	0,10	0,35	0,00
5	0,10	0,15	0,00	0,20	0,45	0,35
7	0,00	0,05	0,05	0,00	0,10	0,80
9	0,05	0,00	0,00	0,05	0,10	0,90
p_x	0,20	0,25	0,20	0,35		1,00
F_x	0,00	0,20	0,45	0,65	1,00	

Prawdopodobieństwa warunkowe obliczamy z definicji I.3.5:

$$\text{a) } P(X=10|Y=7) = \frac{P(X=10 \wedge Y=7)}{P(Y=7)} = \frac{0}{0,10} = 0,$$

$$\text{b) } P((Y=5 \vee Y=2)|X=4) = \frac{P((Y=5 \vee Y=2) \wedge X=4)}{P(X=4)} = \frac{0,00 + 0,15}{0,20} = 0,75. \quad \blacksquare$$

I.3.2. Dwuwymiarowa zmienna losowa ciągła

Definicja I.3.6. Z dwuwymiarową zmienną losową ciągłą (X, Y) mamy do czynienia, jeżeli istnieje nieujemna funkcja $f(x, y)$ taka, że dystrybuantę zmiennej losowej (X, Y) można wyrazić za pomocą następującej całki:

$$F(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x, y) dy dx \quad \text{dla} \quad (x_0, y_0) \in \mathbb{R}^2.$$

Funkcję f nazywamy **gęstością rozkładu prawdopodobieństwa** zmiennej (X, Y) . Podobnie jak dla zmiennej jednowymiarowej – rozkład prawdopodobieństwa dla dwuwymiarowej zmiennej losowej ciągłej może być określony za pomocą funkcji gęstości f lub dystrybuanty F .

Własności dwuwymiarowej zmiennej losowej ciągłej:

$$(1) \quad f \text{ ciągła w } (x, y) \Rightarrow \frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y), \text{ czyli w punktach ciągłości funkcji gęstości jej}$$

wartości równają się pochodnej dystrybuanty;

$$(2) \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1 \quad (I.3.3)$$

$$(3) \quad P(X \in [x_1, x_2] \wedge Y \in [y_1, y_2]) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y) dx dy.$$

I.3.2.1. Rozkłady brzegowe zmiennych losowych ciągłych

Definicja I.3.7. Rozkład brzegowy zmiennej X w rozkładzie dwuwymiarowej zmiennej losowej ciągłej (X, Y) określony jest przez funkcję gęstości zdefiniowaną następująco:

$$f_x(x) = \int_{-\infty}^{+\infty} f(x, y) dy.$$

Analogicznie definiujemy funkcję gęstości dla rozkładu brzegowego zmiennej Y :

$$f_y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Zauważmy, że spełnione są warunki rozkładu:

$$\int_{-\infty}^{+\infty} f_x(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dy dx = 1 \quad \text{oraz} \quad \int_{-\infty}^{+\infty} f_y(y) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1.$$

Dystrybuanty rozkładów brzegowych zmiennej ciągłej określamy następująco:

$$F_x(x_0) = \int_{-\infty}^{x_0} f_x(x) dx, \quad F_y(y_0) = \int_{-\infty}^{y_0} f_y(y) dy \quad (I.3.4)$$

I.3.2.2. Rozkłady warunkowe zmiennych losowych ciągłych

Definicja I.3.8. Rozkład warunkowy zmiennej losowej X , pod warunkiem, że zmienna Y przyjmuje ustalone wartości, jest zdefiniowany następującą funkcją gęstości:

$$f(x|y) = \frac{f(x, y)}{f_y(y)}.$$

Podobnie definiujemy funkcję gęstości dla rozkładu warunkowego zmiennej losowej Y , pod warunkiem, że zmienna X przyjmuje ustalone wartości:

$$f(y|x) = \frac{f(x, y)}{f_x(x)}.$$

Czytelnikowi pozostawiam sprawdzenie, że funkcje z definicji I.3.8. mają własność (2) funkcji gęstości, podaną w podrozdziale I.3.2 (dowód [8], s. 167).

I.3.3. Niezależność zmiennych losowych

Pojęcie niezależności zmiennych losowych jest jednym z ważniejszych pojęć rachunku prawdopodobieństwa. Nawiązuje ono do pojęcia niezależności zdarzeń, wyjaśnionego w definicji I.1.4.

Definicja I.3.9. Zmienne losowe X i Y są niezależne, jeżeli dla dowolnych ich wartości, zdarzenia losowe im odpowiadające są niezależne.

Można sformułować następujące warunki równoważne niezależności zmiennych losowych:

- dla zmiennych dyskretnych:

$$P(X = x_i | Y = y_j) = P(X = x_i) \quad \text{oraz} \quad P(Y = y_j | X = x_i) = P(Y = y_j) \quad (\text{I.3.5a})$$

- dla zmiennych ciągłych:

$$f(x|y) = f_x(x) \quad \text{oraz} \quad f(y|x) = f_y(y) \quad (\text{I.3.5b})$$

Podstawiając wyrażenia (I.3.5a), (I.3.5b) odpowiednio do wyrażeń w definicjach (I.3.5) i (I.3.8), otrzymujemy kolejną postać warunków na niezależność zmiennych losowych:

- dla zmiennych dyskretnych:

$$P_{ij} = P_{i \cdot} \cdot P_{\cdot j} \quad (\text{I.3.6a})$$

- dla zmiennych ciągłych:

$$f(x, y) = f_x(x) \cdot f_y(y) \quad (\text{I.3.6b})$$

I.3.4. Charakterystyki liczbowe zmiennej losowej dwuwymiarowej

Składowe X, Y dwuwymiarowej zmiennej losowej możemy obłożyć dowolną funkcją g , dla której wartość przeciętną definiujemy następująco:

Definicja I.3.10. Wartość przeciętna funkcji g dwuwymiarowej zmiennej losowej (X, Y) , to wynik następującego wyrażenia:

$$E[g(X, Y)] = \begin{cases} \sum_i \sum_j g(x_i, y_j) \cdot p_{ij} & \text{dla zmiennej } (X, Y) \text{ dyskretnej,} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \cdot f(x, y) dx dy & \text{dla zmiennej } (X, Y) \text{ ciągłej.} \end{cases}$$

W nawiązaniu do poprzedniego podrozdziału i omówionego w nim zagadnienia niezależności zmiennych losowych, Czytelnik może sam łatwo udowodnić twierdzenie I.3.1, korzystając z definicji I.3.9. oraz warunków (I.3.6a) i (I.3.6b).

Twierdzenie I.3.1. Jeśli zmienne losowe X i Y są niezależne oraz istnieją ich wartości przeciętne $E(X)$ i $E(Y)$, to spełnione jest równanie: $E(X \cdot Y) = E(X) \cdot E(Y)$.

Uwaga. Twierdzenie odwrotne nie jest prawdziwe.

Podobnie jak dla zmiennej losowej jednowymiarowej wartość przeciętna jest podstawą wyliczenia tzw. momentów mieszanych dwuwymiarowej zmiennej losowej.

Definicja I.3.11. Momentem zwykłym mieszanym rzędu $r+s$ dwuwymiarowej zmiennej losowej (X, Y) nazywamy liczbę $E(X^r \cdot Y^s)$. Natomiast **moment centralny mieszany rzędu $r+s$** to liczba $E[(X - E(X))^r \cdot (Y - E(Y))^s]$. Zakładamy, że $r, s \in N$.

Korzystając z definicji I.3.10 wartości przeciętnej funkcji dwuwymiarowej zmiennej losowej, można napisać wzory na momenty zwykłe i centralne dowolnego rzędu $r+s$ dla zmiennej dyskretnej i zmiennej ciągłej. Zawiera je tabela I.3.6.

Tabela I.3.6

Momenty zwykłe i centralne dwuwymiarowej zmiennej losowej

Zmienna dyskretna	Zmienna ciągła
Moment zwykły mieszany $\alpha_{rs} = E(X^r Y^s)$	
$\alpha_{rs} = \sum_i \sum_j x_i^r y_j^s \cdot p_{ij}$	$\alpha_{rs} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^r y^s f(x, y) dx dy$
Moment centralny mieszany $\mu_{rs} = E[(X - E(X))^r \cdot (Y - E(Y))^s]$	
$\mu_{rs} = \sum_i \sum_j (x_i - E(X))^r (y_j - E(Y))^s p_{ij}$	$\mu_{rs} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))^r (y - E(Y))^s f(x, y) dx dy$

Zauważmy, że:

– dla $s = 0$ otrzymujemy:

- moment zwykły rzędu r w rozkładzie brzegowym zmiennej X :

$$\alpha_{r0} = E(X^r \cdot Y^0) = E(X^r) = \alpha_r(X);$$

- moment centralny rzędu r w rozkładzie brzegowym zmiennej X :

$$\mu_{r0} = E[(X - E(X))^r \cdot (Y - E(Y))^0] = E[(X - E(X))^r] = \mu_r(X);$$

– dla $r = 0$ otrzymujemy:

- moment zwykły rzędu s w rozkładzie brzegowym zmiennej Y :

$$\alpha_{0s} = E(X^0 \cdot Y^s) = E(Y^s) = \alpha_s(Y);$$

- moment centralny rzędu s w rozkładzie brzegowym zmiennej Y :

$$\mu_{0s} = E[(X - E(X))^0 \cdot (Y - E(Y))^s] = E[(Y - E(Y))^s] = \mu_s(Y);$$

– dla $r = 1$ oraz $s = 0$: $\alpha_{10} = E(X^1 \cdot Y^0) = E(X)$ – wartość przeciętna zmiennej X ;

– dla $r = 0$ oraz $s = 1$: $\alpha_{01} = E(X^0 \cdot Y^1) = E(Y)$ – wartość przeciętna zmiennej Y ;

– dla $r = 2$ oraz $s = 0$: $\mu_{20} = E[(X - E(X))^2 \cdot (Y - E(Y))^0] = V(X)$ – wariancja zmiennej X ;

– dla $r = 0$ oraz $s = 2$: $\mu_{02} = E[(X - E(X))^0 \cdot (Y - E(Y))^2] = V(Y)$ – wariancja zmiennej Y .

Warto jeszcze zwrócić uwagę na inną charakterystykę dwuwymiarowej zmiennej losowej:

kowariancja zmiennych X i Y jest to moment centralny mieszany rzędu 1+1

$$\text{dla } r = 1 \text{ oraz } s = 1: \quad \text{cov}(X, Y) = \mu_{11} = E[(X - E(X))^1 \cdot (Y - E(Y))^1] \quad (\text{I.3.7})$$

Z twierdzenia I.3.1. oraz z własności wartości oczekiwanej (I.2.4) wynika, że dla zmiennych losowych niezależnych X i Y zachodzi: $\text{cov}(X, Y) = 0$.

Jak zostało pokazane w przykładzie I.2.6, momenty centralne można wyrazić za pomocą momentów zwykłych. Pozostawiam Czytelnikowi do samodzielnego wyprowadzenia następujące wzory:

$$\mu_{20} = \alpha_{20} - \alpha_{10}^2, \quad \text{czyli} \quad V(X) = E(X^2) - E^2(X)$$

$$\mu_{02} = \alpha_{02} - \alpha_{01}^2, \quad \text{czyli} \quad V(Y) = E(Y^2) - E^2(Y) \quad (\text{I.3.8})$$

$$\mu_{11} = \alpha_{11} - \alpha_{10}\alpha_{01}, \quad \text{czyli} \quad \text{cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

Odpowiednikiem wartości przeciętnej jednowymiarowej zmiennej losowej $E(X)$, w przypadku dwuwymiarowym jest **wektor wartości przeciętnych** $[E(X), E(Y)]$, wyznaczonych dla rozkładów brzegowych. Natomiast wariancji jednowymiarowej zmiennej $V(X)$ odpowiada macierz momentów centralnych drugiego rzędu, zwana **macierzą kowariancji** $Cov(X, Y)$.

$$Cov(X, Y) = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} = \begin{bmatrix} V(X) & cov(X, Y) \\ cov(X, Y) & V(Y) \end{bmatrix} \quad (I.3.9)$$

Z definicji macierz $Cov(X, Y)$ jest zawsze symetryczna, a jej wyznacznik jest zawsze nieujemny, czyli:

$$\det(Cov(X, Y)) = V(X)V(Y) - cov^2(X, Y) \geq 0 \quad (I.3.10)$$

PRZYKŁAD I.3.3

Dla zmiennej losowej (X, Y) z przykładu I.3.1 wyznaczyć momenty zwykłe mieszane rzędu 3 oraz centralne mieszane rzędu co najwyżej 2.

Rozwiązanie

Korzystając z definicji I.3.11 wiemy, że chodzi o wyznaczenie momentów zwykłych α_{rs} takich, że $r + s = 3$ oraz centralnych μ_{rs} takich, że $r + s \leq 2$. W obliczeniach wykorzystamy ze wzorów w tabeli I.3.6.

Momenty zwykłe:

$$\alpha_{30} = \sum_i \sum_j x_i^3 y_j^0 \cdot p_{ij} = \sum_i x_i^3 \cdot p_{i\cdot} = 1^3 \cdot 0,20 + 2^3 \cdot 0,25 + 4^3 \cdot 0,20 + 10^3 \cdot 0,35 = 365,$$

$$\alpha_{03} = \sum_i \sum_j x_i^0 y_j^3 \cdot p_{ij} = \sum_j y_j^3 \cdot p_{\cdot j} = 2^3 \cdot 0,35 + 5^3 \cdot 0,45 + 7^3 \cdot 0,10 + 9^3 \cdot 0,10 = 166,25,$$

$$\begin{aligned} \alpha_{12} &= \sum_i \sum_j x_i^1 y_j^2 \cdot p_{ij} = 1 \cdot (2^2 \cdot 0,05 + 5^2 \cdot 0,10 + 7^2 \cdot 0,00 + 9^2 \cdot 0,05) + \\ &+ 2 \cdot (2^2 \cdot 0,05 + 5^2 \cdot 0,15 + 7^2 \cdot 0,05 + 9^2 \cdot 0,00) + \\ &+ 4 \cdot (2^2 \cdot 0,15 + 5^2 \cdot 0,00 + 7^2 \cdot 0,05 + 9^2 \cdot 0,00) + \\ &+ 10 \cdot (2^2 \cdot 0,10 + 5^2 \cdot 0,20 + 7^2 \cdot 0,00 + 9^2 \cdot 0,05) = 6,75 + 12,8 + 12,2 + 94,5 = 126,25, \end{aligned}$$

$$\begin{aligned} \alpha_{21} &= \sum_i \sum_j x_i^2 y_j^1 \cdot p_{ij} = 1^2 \cdot (2 \cdot 0,05 + 5 \cdot 0,10 + 7 \cdot 0,00 + 9 \cdot 0,05) + \\ &+ 2^2 \cdot (2 \cdot 0,05 + 5 \cdot 0,15 + 7 \cdot 0,05 + 9 \cdot 0,00) + 4^2 \cdot (2 \cdot 0,15 + 5 \cdot 0,00 + 7 \cdot 0,05 + 9 \cdot 0,00) + \\ &+ 10^2 \cdot (2 \cdot 0,10 + 5 \cdot 0,20 + 7 \cdot 0,00 + 9 \cdot 0,05) = 1,05 + 4,8 + 10,4 + 165 = 181,25. \end{aligned}$$

Momenty centralne:

Wiemy, że moment centralny mieszany μ_{20} rzędu dwa-zero jest wariancją zmiennej X . Aby wykorzystać uproszczony sposób jej wyznaczania ze wzorów (I.3.8), obliczamy najpierw, na podstawie rozkładu brzegowego, potrzebne momenty zwykłe zmiennej jednowymiarowej X :

$$E(X) = 1 \cdot 0,20 + 2 \cdot 0,25 + 4 \cdot 0,20 + 10 \cdot 0,35 = 0,20 + 0,50 + 0,80 + 3,50 = 5,$$

$$E(X^2) = 1^2 \cdot 0,20 + 2^2 \cdot 0,25 + 4^2 \cdot 0,20 + 10^2 \cdot 0,35 = 0,2 + 1 + 3,2 + 35 = 39,4.$$

Wyliczone momenty zwykłe podstawiamy do wzoru (I.3.8):

$$\mu_{20} = V(X) = E(X^2) - E^2(X) = 39,4 - 5^2 = 14,4.$$

Analogicznie obliczymy moment centralny mieszany μ_{02} rzędu zero-dwa, który jest wariancją zmiennej Y . Najpierw wyliczamy potrzebne momenty zwykłe:

$$E(Y) = 2 \cdot 0,35 + 5 \cdot 0,45 + 7 \cdot 0,10 + 9 \cdot 0,10 = 0,7 + 2,25 + 0,7 + 0,9 = 4,55,$$

$$E(Y^2) = 2^2 \cdot 0,35 + 5^2 \cdot 0,45 + 7^2 \cdot 0,10 + 9^2 \cdot 0,10 = 1,4 + 11,25 + 4,9 + 8,1 = 25,65.$$

Stąd:

$$\mu_{02} = V(Y) = E(Y^2) - E^2(Y) = 25,65 - (4,55)^2 = 4,9475.$$

Na podstawie wzorów (I.3.8) obliczymy również mieszany moment centralny μ_{11} rzędu jeden-jeden, czyli kowariancję. Zaczniemy znowu od potrzebnego momentu zwykłego α_{11} :

$$\begin{aligned} \alpha_{11} = E(X \cdot Y) &= 1 \cdot (2 \cdot 0,05 + 5 \cdot 0,10 + 7 \cdot 0,00 + 9 \cdot 0,05) + \\ &+ 2 \cdot (2 \cdot 0,05 + 5 \cdot 0,15 + 7 \cdot 0,05 + 9 \cdot 0,00) + \\ &+ 4 \cdot (2 \cdot 0,15 + 5 \cdot 0,00 + 7 \cdot 0,05 + 9 \cdot 0,00) + \\ &+ 10 \cdot (2 \cdot 0,10 + 5 \cdot 0,20 + 7 \cdot 0,00 + 9 \cdot 0,05) = 22,55, \end{aligned}$$

$$\begin{aligned} \mu_{11} &= \sum_i \sum_j (x_i - E(X))^1 (y_j - E(Y))^1 p_{ij} = \\ &= cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 22,55 - 5 \cdot 4,55 = -0,2 \end{aligned}$$

Nierówność $r + s \leq 2$, będącą warunkiem dla rzędu momentów centralnych, spełniają również następujące pary $1 + 0$, $0 + 1$ i $0 + 0$:

$$\begin{aligned} \mu_{10} &= \sum_i \sum_j (x_i - E(X))^1 (y_j - E(Y))^0 p_{ij} = \\ &= \sum_i (x_i - E(X)) p_i = \sum_i x_i \cdot p_i - E(X) \cdot \sum_i p_i = \dots \end{aligned}$$

Ponieważ pierwsza suma odpowiada definicji (I.2.5) wartości przeciętnej, a druga ze względu na warunki rozkładu zmiennej losowej dyskretnej (definicja I.2.3) wynosi 1, otrzymujemy:

$$\dots = E(X) - E(X) \cdot 1 = 0.$$

Analogicznie uzyskamy zero dla momentu centralnego μ_{01} .

Ostatni z momentów przyjmuje z definicji I.3.11 uproszczoną formę:

$$\mu_{00} = E[(X - E(X))^0 \cdot (Y - E(Y))^0] = E(1) = 1. \quad \blacksquare$$

I.3.5. Współczynnik korelacji

Współczynnik korelacji liniowej jest ważną charakterystyką dwuwymiarowej zmiennej losowej. Jest to miara liniowej zależności między dwiema zmiennymi losowymi X i Y , zdefiniowana następująco:

$$r = \frac{\text{cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} \quad (\text{I.3.11})$$

Ze względu na własność (I.3.10) macierzy kowariancji – współczynnik korelacji przyjmuje zawsze wartości z przedziału $[-1; 1]$. Jeżeli r wynosi dokładnie 1 lub -1 , jest to równoznaczne z faktem pełnej zależności liniowej między zmiennymi. Wówczas można jedną zmienną przedstawić jako kombinację liniową drugiej zmiennej, czyli w postaci: $Y = a \cdot X + b$ lub $X = a \cdot Y + b$. Jeżeli natomiast $r = 0 \Leftrightarrow \text{cov}(X, Y) = 0$, to brak jest zależności liniowej, a zmienne X, Y nazywamy nieskorelowanymi. Pośrednie wartości współczynnika korelacji informują o stopniu skorelowania pary zmiennych. Ocena nasilenia skorelowania zmiennych na podstawie wartości współczynnika korelacji jest zależna od rodzaju badanego zagadnienia, ale umownie można przyjąć następującą przybliżoną klasyfikację:

$$\begin{aligned} |r| \in (0; 0,2] & - \text{korelacja bardzo słaba} \\ |r| \in (0,2; 0,4] & - \text{korelacja słaba} \\ |r| \in (0,4; 0,6] & - \text{korelacja średnia} \\ |r| \in (0,6; 0,8] & - \text{korelacja mocna} \\ |r| \in (0,8; 1] & - \text{korelacja bardzo mocna} \end{aligned} \quad (\text{I.3.12})$$

Ze względu na fakt, że zależność liniowa jest jednym z wielu możliwych rodzajów zależności, zachodzą następujące implikacje:

X, Y – skorelowane $\Rightarrow X, Y$ – zależne oraz X, Y – niezależne $\Rightarrow X, Y$ – nieskorelowane.

Zmienne mogą być niezależne liniowo, ale zależne w ustalony sposób nieliniowy, np.:

$$Y = a \cdot X^2 + b \cdot X + c.$$

I.3.6. Proste regresji drugiego rodzaju

Definicja I.3.12. Prosta regresji drugiego rodzaju zmiennej losowej Y względem zmiennej losowej X , to prosta o równaniu $Y = a \cdot X + b$, dla której współczynniki a i b są tak dobrane, by spełniony był warunek:

$$E[Y - (aX + b)]^2 = \min.$$

Powyższy warunek oznacza minimalizację odchylenia średniokwadratowego zmiennej Y od zmiennej $aX + b$. Można dowieść (patrz [5]), że dla dowolnej dwuwymiarowej zmiennej losowej (X, Y) , o skończonych wariancjach w rozkładach brzegowych $V(X)$ i $V(Y)$, istnieje dokładnie jedna prosta, o której mowa w definicji I.3.12. Jej parametry określone są następującymi wzorami:

$$a = r \cdot \frac{\sigma(Y)}{\sigma(X)} \quad (\text{I.3.13a})$$

$$b = E(Y) - r \cdot \frac{\sigma(Y)}{\sigma(X)} \cdot E(X) \quad (\text{I.3.13b})$$

Po podstawieniu równania (I.3.13a) do (I.3.13b) otrzymujemy: $b = E(Y) - a \cdot E(X)$. Natomiast po wstawieniu a i b do równania prostej regresji otrzymujemy:

$$\frac{Y - E(Y)}{\sigma(Y)} = r \cdot \frac{X - E(X)}{\sigma(X)} \quad (\text{I.3.14})$$

Analogicznie uzyskuje się równanie linii regresji zmiennej X względem zmiennej Y :

$$\frac{X - E(X)}{\sigma(X)} = r \cdot \frac{Y - E(Y)}{\sigma(Y)} \quad (\text{I.3.15})$$

Metodę wyznaczania prostych regresji drugiego rodzaju, minimalizującą średnie odchylenie kwadratowe, nazywamy **metodą najmniejszych kwadratów** (MNK). Zastosowanie MNK w tym zagadnieniu zostanie szerzej omówione w podrozdziale II.4.2.1.

Proste regresji $Y(X)$ i $X(Y)$ pokrywają się, gdy $|r| = 1$.

PRZYKŁAD I.3.4

Tabela I.3.7 przedstawia rozkład dwuwymiarowej zmiennej losowej (X, Y) , w której X oznacza wydajność pracy liczoną w tysiącach sztuk, a Y oznacza czas dojazdu do pracy w kwadransach.

Tabela I.3.7

Rozkład dwuwymiarowej zmiennej losowej dyskretnej (X, Y)

$Y \backslash X$	4	5	6
1	0,0	0,0	0,1
2	0,1	0,2	0,1
3	0,3	0,1	0,1

Udowodnić, że zmienne losowe X i Y są zależne. Wykonać prognozę czasu dojazdu do pracy pracownika, którego wydajność wynosi 5,5 tysiąca sztuk oraz prognozę wydajności pracownika, któremu droga do pracy zajmuje 2,5 kwadransu.

Rozwiązanie

Chcąc potwierdzić zależność zmiennych losowych, wystarczy znaleźć choć jedną parę (X_i, Y_j) , dla której nie jest spełniony warunek niezależności (I.3.6a): $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$, czyli wystarczy znaleźć tzw. kontrprzykład do warunku równoważnego definicji I.3.9 – niezależności zmiennych losowych:

$$P(X=5 \wedge Y=1) = 0,$$

$$P(X=5) = 0,0 + 0,2 + 0,1 = 0,3,$$

$$P(Y=1) = 0,0 + 0,0 + 0,1 = 0,1,$$

$$P(X=5) \cdot P(Y=1) = 0,3 \cdot 0,1 = 0,03,$$

Widać, że: $P(X=5 \wedge Y=1) \neq P(X=5) \cdot P(Y=1)$, czyli zmienne losowe X i Y nie są niezależne.

W celu wykonania żądanych prognoz, wyznaczmy linie regresji drugiego rodzaju Y względem X oraz X względem Y . Najpierw jednak należy określić rozkłady brzegowe zmiennych p_x oraz p_y (tab. I.3.8).

Tabela I.3.8

Rozkład zmiennej losowej (X, Y) wraz z rozkładami brzegowymi

$Y \backslash X$	4	5	6	p_y
1	0,0	0,0	0,1	0,1
2	0,1	0,2	0,1	0,4
3	0,3	0,1	0,1	0,5
p_x	0,4	0,3	0,3	1,0

Na podstawie rozkładów brzegowych wyznaczamy potrzebne charakterystyki, tzn. wartości oczekiwane oraz wariancje (ze wzorów (I.3.8)) obu zmiennych losowych $E(X)$, $E(Y)$, $V(X)$, $V(Y)$:

$$E(X) = 4 \cdot 0,4 + 5 \cdot 0,3 + 6 \cdot 0,3 = 1,6 + 1,5 + 1,8 = 4,9,$$

$$E(Y) = 1 \cdot 0,1 + 2 \cdot 0,4 + 3 \cdot 0,5 = 0,1 + 0,8 + 1,5 = 2,4,$$

$$E(X^2) = 4^2 \cdot 0,4 + 5^2 \cdot 0,3 + 6^2 \cdot 0,3 = 6,4 + 7,5 + 10,8 = 24,7,$$

$$E(Y^2) = 1^2 \cdot 0,1 + 2^2 \cdot 0,4 + 3^2 \cdot 0,5 = 0,1 + 1,6 + 4,5 = 6,2,$$

$$V(X) = E(X^2) - E^2(X) = 24,7 - (4,9)^2 = 0,69,$$

$$V(Y) = E(Y^2) - E^2(Y) = 6,2 - (2,4)^2 = 0,44,$$

a z rozkładu zmiennej dwuwymiarowej obliczamy moment zwykły mieszany rzędu 1+1, czyli następującą wartość oczekiwaną:

$$\begin{aligned}\alpha_{11} &= E(X^1 \cdot Y^1) = E(XY) = \sum_{i=1}^3 \sum_{j=1}^3 x_i \cdot y_j \cdot p_{ij} = \\ &= 4 \cdot 1 \cdot 0 + 4 \cdot 2 \cdot 0,1 + 4 \cdot 3 \cdot 0,3 + 5 \cdot 1 \cdot 0 + 5 \cdot 2 \cdot 0,2 + 5 \cdot 3 \cdot 0,1 + 6 \cdot 1 \cdot 0,1 + 6 \cdot 2 \cdot 0,1 + 6 \cdot 3 \cdot 0,1 = \\ &= 0 + 0,8 + 3,6 + 0 + 2 + 1,5 + 0,6 + 1,2 + 1,8 = 11,5,\end{aligned}$$

Mając powyższe wyniki, możemy wyznaczyć kowariancję między zmiennymi $cov(X, Y)$, a z niej współczynnik korelacji $r(X, Y)$. Ze wzoru (I.3.8) mamy:

$$cov(X, Y) = E(XY) - E(X)E(Y) = 11,5 - 4,9 \cdot 2,4 = -0,26.$$

Ujemna kowariancja pociągnie za sobą ujemną wartość współczynnika korelacji r . Oznacza to spadek wartości jednej zmiennej przy wzroście wartości drugiej zmiennej. Siłę korelacji oceniamy na podstawie wartości r , obliczonej ze wzoru (I.3.11):

$$r = \frac{cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{-0,26}{\sqrt{0,69 \cdot 0,44}} = -0,47187.$$

Korelacja między zmiennymi X i Y jest na poziomie średnim.

Można teraz przystąpić do wyznaczenia linii regresji według wzorów (I.3.14) oraz (I.3.15). Pierwszy z nich po przekształceniu przyjmuje postać:

$$\begin{aligned}Y &= r \cdot \frac{\sigma(Y)}{\sigma(X)} \cdot X - r \cdot \frac{\sigma(Y)}{\sigma(X)} \cdot E(X) + E(Y) = \\ &= \frac{cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} \cdot \sqrt{\frac{V(Y)}{V(X)}} \cdot X - \frac{cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} \cdot \sqrt{\frac{V(Y)}{V(X)}} \cdot E(X) + E(Y) = \\ &= \frac{cov(X, Y)}{V(X)} \cdot X - \frac{cov(X, Y)}{V(X)} \cdot E(X) + E(Y) = \\ &= \frac{-0,26}{0,69} \cdot X - \frac{-0,26}{0,69} \cdot 4,9 + 2,4 = -0,3768 \cdot X + 4,2464.\end{aligned}$$

Zatem linia regresji Y względem X jest postaci:

$$Y = -0,377 \cdot X + 4,246.$$

Analogicznie uzyskuje się równanie linii regresji zmiennej X względem zmiennej Y :

$$\begin{aligned}X &= r \cdot \frac{\sigma(X)}{\sigma(Y)} \cdot Y - r \cdot \frac{\sigma(X)}{\sigma(Y)} \cdot E(Y) + E(X) = \frac{cov(X, Y)}{V(Y)} \cdot Y - \frac{cov(X, Y)}{V(Y)} \cdot E(Y) + E(X) = \\ &= \frac{-0,26}{0,44} \cdot Y - \frac{-0,26}{0,44} \cdot 2,4 + 4,9 = -0,5909 \cdot Y + 6,3182.\end{aligned}$$

Zatem linia regresji X względem Y jest postaci:

$$X = -0,591 \cdot Y + 6,318.$$

Możemy teraz wykonać prognozy zadane w treści zadania. Chcąc wyznaczyć przeciętny czas dojazdu do pracy pracownika, którego wydajność wynosi 5,5 tysiąca sztuk, korzystamy z linii regresji Y względem X , ponieważ prognozujemy zmienną Y : $Y = -0,377 \cdot 5,5 + 4,246 = 2,17$, czyli pracownik o takiej wydajności średnio potrzebuje nieco ponad dwa kwadransy na dotarcie do pracy. Podobnie, w celu oszacowania wydajności pracownika, który dwa i pół kwadransu spędza w drodze do pracy, korzystamy z linii regresji X względem Y :

$$X = -0,591 \cdot 2,5 + 6,318 = 4,84,$$

czyli wydajność takiego pracownika wynosi niecałe 5 tysięcy sztuk.

■

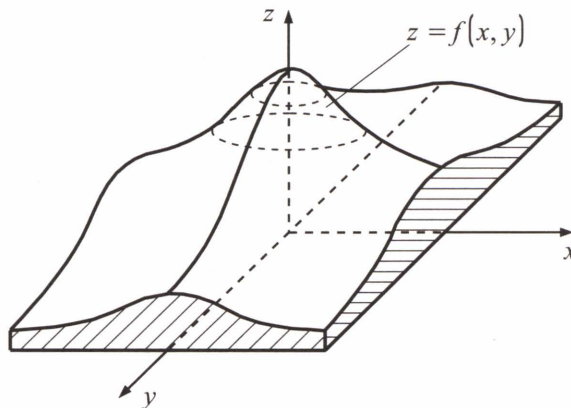
I.3.7. Dwuwymiarowy rozkład normalny

Szczególnym przykładem rozkładu dwuwymiarowej zmiennej losowej jest dwuwymiarowy rozkład normalny. Jest on bardzo ważny ze względu na liczne zastosowania. Funkcja gęstości dla tego rozkładu jest następującej postaci:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} \cdot \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2r\frac{(x-\mu_x)(y-\mu_y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \quad (\text{I.3.16})$$

dla $(x, y) \in \mathbb{R}^2$.

Wykres funkcji (I.3.16) dla $E(X) = E(Y) = 0$ przedstawia rysunek I.3.1.



Rys. I.3.1. Gęstość dwuwymiarowego rozkładu normalnego o wartości przeciętnej równej zero

Przekrój poziomy powyższego wykresu na zadanej wysokości z_0 jest elipsą. Jest to tzw. **elipsa stałej gęstości prawdopodobieństwa** o równaniu:

$$\frac{(x-E(X))^2}{V(X)} - 2r \frac{(x-E(X))(y-E(Y))}{\sigma_X \sigma_Y} + \frac{(y-E(Y))^2}{V(Y)} = c^2 = \text{const} \quad (\text{I.3.17})$$

gdzie $c^2 = -2(1-r^2) \ln(2\pi\sigma_X\sigma_Y\sqrt{1-r^2} \cdot z_0)$.

Dla $c = 1$, parametry elipsy wyrażają się następującymi wzorami:

$$\left. \begin{aligned} \sigma_{\max}^2 \\ \sigma_{\min}^2 \end{aligned} \right\} = \frac{V(X)+V(Y)}{2} \pm \sqrt{\left(\frac{V(X)-V(Y)}{2}\right)^2 + \text{cov}^2(X,Y)} \quad (\text{I.3.18})$$

$$\beta = \frac{1}{2} \arctg \frac{2 \cdot \text{cov}(X,Y)}{V(X)-V(Y)}$$

gdzie:

$\sigma_{\max}, \sigma_{\min}$ – półosie elipsy,

β – kąt skręcenia elipsy (azymut dłuższej półosi σ_{\max}).

Jeżeli zmienne są nieskorelowane ($r = 0$), to osie elipsy są równoległe do osi układu współrzędnych X, Y , a jej równanie (I.3.17) przyjmuje postać:

$$\frac{(x-E(X))^2}{V(X)} + \frac{(y-E(Y))^2}{V(Y)} = c^2 = \text{const}.$$

Zatem wraz ze zmniejszaniem się zależności korelacyjnej pomiędzy zmiennymi X, Y , kierunek osi elipsy zbliża się do kierunku osi współrzędnych.

Własności dwuwymiarowego rozkładu normalnego:

(1) dla dwuwymiarowego rozkładu normalnego zachodzi równoważność:

X, Y – nieskorelowane $\Leftrightarrow X, Y$ – niezależne;

(2) rozkłady brzegowe w dwuwymiarowym rozkładzie normalnym są jednowymiarowymi rozkładami normalnymi;

(3) funkcja gęstości osiąga maksimum dla punktu o współrzędnych $(x, y) = (E(X), E(Y))$; jest to również środek każdej elipsy stałej gęstości;

(4) zbiór punktów w przestrzeni R^2 , spełniających równania (I.3.14) lub (I.3.15):

$$\frac{Y - \mu_Y}{\sigma_Y} = r \cdot \frac{X - \mu_X}{\sigma_X} \quad \text{lub} \quad \frac{X - \mu_X}{\sigma_X} = r \cdot \frac{Y - \mu_Y}{\sigma_Y}$$

to tzw. **linie regresji pierwszego rodzaju** zmiennej Y względem zmiennej X i zmiennej X względem zmiennej Y ; dla rozkładu normalnego pokrywają się one z prostymi regresji II rodzaju;

- (5) odcinki łączące punkty styczności elipsy stałej gęstości i prostokąta stycznego do tej elipsy, o bokach równoległych do osi układu współrzędnych, należą do linii regresji pierwszego rodzaju; na rysunku I.3.2 prosta AB jest linią regresji Y względem X , a prosta CD – linią regresji X względem Y ;
- (6) dla $r = 0$, półosie elipsy są równoległe do osi układu współrzędnych;
- (7) z odwrotności macierzy kowariancji

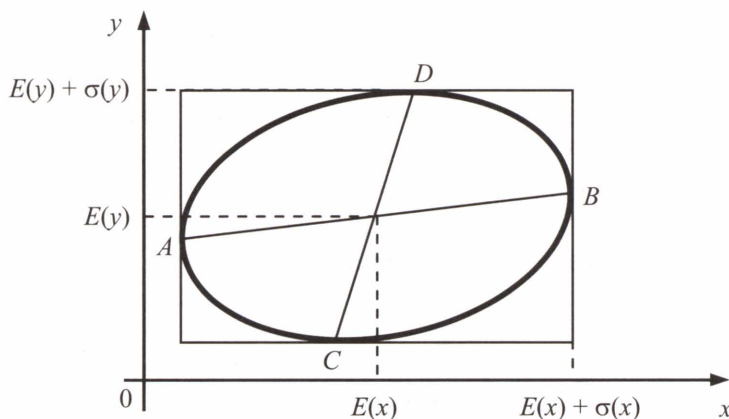
$$\text{Cov}^{-1}(X, Y) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

można odczytać parametry funkcji gęstości:

$$f(x, y) = \frac{1}{2\pi\sqrt{\det \text{Cov}(X, Y)}} \cdot \exp \left\{ -\frac{1}{2} \left[b_{11} \cdot (x - E(X))^2 + b_{12} \cdot (x - E(X))(y - E(Y)) + b_{22} \cdot (y - E(Y))^2 \right] \right\}$$

oraz linii regresji pierwszego rodzaju zmiennej X względem zmiennej Y :

$$b_{11} \cdot (x - E(X)) = -b_{12} (y - E(Y)).$$



Rys. I.3.2. Elipsa stałej gęstości prawdopodobieństwa (rzut przekroju poprzecznego wykresu gęstości dwuwymiarowego rozkładu normalnego na płaszczyznę XY)

II. Elementy statystyki matematycznej

II.1. Populacja, próba, cecha

Słowo „statystyka” pochodzi od włoskiego słowa *stato*, czyli państwo. Dzisiaj statystyki nie ogranicza się do informacji dotyczących spraw państwowych. Przedmiotem badań statystycznych może być niemal każda dziedzina działalności człowieka.

Rozróżniamy tzw.:

- statystykę opisową
- statystykę matematyczną.

Pierwsza z nich ogranicza się do opisu wyników uzyskanych w doświadczeniu i zgromadzonych w próbie, bez odnoszenia ich do całej populacji. Natomiast druga, zajmuje się wnioskowaniem nt. badanej cechy w całej populacji (tzw. **populacja generalna**), na podstawie wyników w tzw. **próbie losowej (statystycznej)**.

Statystyka matematyczna zajmuje się analizą zjawisk masowych, a narzędziem analiz są dla niej metody rachunku prawdopodobieństwa. Analizie może podlegać jedna lub więcej cech badanej zbiorowości **populacji generalnej**. Elementami populacji mogą być obiekty podlegające badaniu, albo wartości badanej cechy. Na populacji można przeprowadzić tzw. **badanie kompletne**, uwzględniając w analizach wszystkie elementy populacji. Na przykład pomiar wzrostu i wagi wszystkich uczniów danej szkoły, w określonym czasie, w celu wyciągnięcia wniosków o przeciętnych wartościach tych cech w badanej zbiorowości uczniów. Jednakże najczęściej badanie kompletne jest zbyt czasochłonne, kosztowne lub wręcz niemożliwe, choćby ze względu na jego niszczący charakter (np. badania wytrzymałości materiałów na oddziaływanie czynników zewnętrznych: mechanicznych, chemicznych itp., badanie jakości produktów spożywczych, badanie odporności roślin na czynniki chemiczne itd.), dlatego też w praktyce przeprowadza się analizy na części populacji, czyli na próbie statystycznej.

Badania kompletne dostarczają pełnej informacji o badanej cenie, ale nie są przedmiotem analiz statystycznych, gdyż zadaniem statystyki matematycznej jest wnioskowanie nt. własności populacji na podstawie informacji uzyskanych z próby statystycznej. Dlatego próba powinna być tzw. próbą reprezentatywną w stosunku do populacji generalnej, tzn. rozkład badanej cechy w próbie powinien być zbliżony do rozkładu tej samej cechy w całej

populacji. Uzyskuje się to poprzez losowy wybór elementów z populacji do próby, a taką próbę nazywamy właśnie **próbą losową**. Wnioski statystyczne, ze względu na fakt, iż opierają się na informacji częściowej, można uznać co najwyżej za wiarygodne, a nie bezwzględnie prawdziwe.

Przedmiotem badania statystycznego może być jedna lub więcej cech populacji generalnej.

Badane cechy populacji możemy podzielić na dwie grupy:

- 1) cechy mierzalne, zwane ilościowymi (np. wielkość zanieczyszczeń emitowanych przez zakład, stężenie roztworu chemicznego, wzrost ludzki, temperatura, poziom opadów itd.),
- 2) cechy niemierzalne, zwane jakościowymi (np. płeć, poziom wykształcenia, preferencje konsumentów nt. marki towaru itd.).

Zmiennym ilościowym w sposób naturalny odpowiadają wartości liczbowe. Zmiennym jakościowym są one umownie przypisywane, na podstawie ustalonej wcześniej skali zmienności, poprzez nadanie tzw. rang ich wartościom. W ten sposób cechy niemierzalne można w konkretnych analizach statystycznych traktować jak mierzalne.

Mówiąc o zmiennej losowej w teorii rachunku prawdopodobieństwa, zakłada się znajomość jej rozkładu i na tej podstawie wyznacza się prawdopodobieństwo różnych zdarzeń. Natomiast w statystyce nie zakłada się znajomości rozkładu zmiennej losowej. Wnioskowanie na temat typu rozkładu i różnych jego cech przeprowadza się na podstawie wylosowanych elementów z całej badanej zbiorowości.

II.2. Statystyka opisowa

II.2.1. Szereg rozdzielczy

Definicja II.2.1. Niech x_1, x_2, \dots, x_n będzie n -elementową próbą statystyczną. **Rozstęp** R cechy X w próbie to różnica między największym i najmniejszym elementem z próby:

$$R = \max(x_1, x_2, \dots, x_n) - \min(x_1, x_2, \dots, x_n) \quad (\text{II.2.1})$$

Przy licznej próbie ($n > 30$), w celu ułatwienia analiz, stosuje się podział jej elementów na klasy, ustalając dla każdej klasy jej reprezentanta. Najczęściej jest to środek klasy. Liczbę klas k oraz długość klasy d można orientacyjnie ustalić w następujący sposób:

$$k \approx \sqrt{n}, \quad d \approx R/k \quad (\text{II.2.2})$$

Liczbę wartości zawartych w i -tej klasie nazywamy liczebnością i -tej klasy i oznaczamy n_i . Oczywiście jest, że: $\sum_{i=1}^k n_i = n$.

Definicja II.2.2. Szereg rozdzielczy to zbiór par liczbowych (\bar{x}_i, n_i) , złożonych ze środków kolejnych klas \bar{x}_i oraz ich liczebności n_i . Reprezentantem klasy i jest jej środek \bar{x}_i .

W tabeli II.2.1 pokazano przykładowy rezultat utworzenia szeregu rozdzielczego ze średnich miesięcznych wartości zużycia energii elektrycznej, w prywatnych mieszkaniach budynku wielorodzinnego. Odczyty zużycia energii pogrupowano w klasy i każdej z nich przyporządkowano jej reprezentanta.

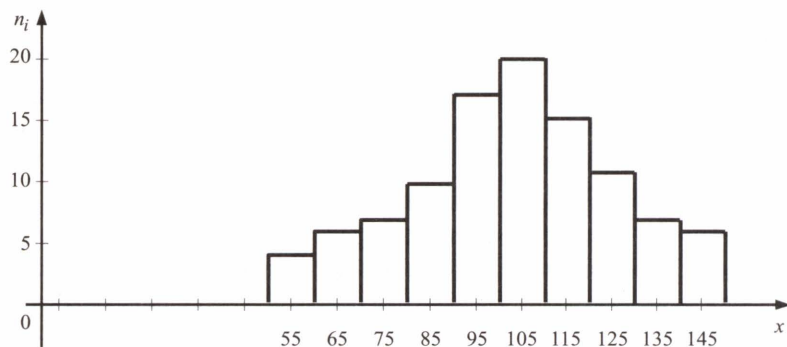
Wszelkie analizy statystyczne prowadzimy na reprezentantach.

Tabela II.2.1

Szereg rozdzielczy dla zmiennej losowej ciągłej X przyjmującej wartości w przedziale [50, 150] [kWh]

Nr klasy i	Klasa zużycie energii [kWh]	Szereg rozdzielczy	
		Środek klasy \bar{x}_i	Liczebność klasy n_i
1	[50÷60)	55	4
2	[60÷70)	65	6
3	[70÷80)	75	7
4	[80÷90)	85	10
5	[90÷100)	95	17
6	[100÷110)	105	20
7	[110÷120)	115	15
8	[120÷130)	125	11
9	[130÷140)	135	7
10	[140÷150)	145	6

Dane z tabeli II.2.1 można graficznie zobrazować w postaci tzw. histogramu (rys. II.2.1).



Rys. II.2.1. Histogram zużycia energii elektrycznej

PRZYKŁAD II.2.1

Dla danych średniego miesięcznego zużycia energii w mieszkaniach pewnego budynku wielorodzinnego (tab. II.2.1) podać wartości dystrybuanty zmiennej losowej X . Liczności n_i odpowiadają liczbie mieszkań zużywających średnio określoną ilość energii. Za pomocą wartości dystrybuanty wyrazić i obliczyć następujące prawdopodobieństwa:

$$P(X > 105), \quad P(100 \leq X < 130), \quad P(80 > X).$$

Obliczyć odchylenie standardowe zmiennej losowej X .

Rozwiązanie

Jako reprezentatywne wartości zmiennej losowej wybieramy środki poszczególnych przedziałów. Mając na uwadze fakt, że mieszkań, z których zebrano dane, jest 103, wyznaczamy rozkład zmiennej X na podstawie tzw. częstościowej definicji prawdopodobieństwa, która została przytoczona w punkcie (7) własności prawdopodobieństwa (I.1.1).

Tabela II.2.2

Rozkład prawdopodobieństwa zmiennej losowej X reprezentującej średnie miesięczne zużycie energii w mieszkaniach ustalonego budynku wielorodzinnego oraz wartości dystrybuanty zmiennej X

\bar{x}_i [kWh]	55	65	75	85	95	105	115	125	135	145	$X > 150$
p_i	4/103	6/103	7/103	10/103	17/103	20/103	15/103	11/103	7/103	6/103	0
$F(x_i)$	0	4/103	10/103	17/103	27/103	44/103	64/103	79/103	90/103	97/103	1

Następnie obliczamy wartości dystrybuanty (tab. II.2.2) i za ich pomocą wyznaczamy żądane prawdopodobieństwa:

$$P(X > 105) = 1 - P(X \leq 105) = 1 - F(105^+) = 1 - F(115) = 1 - \frac{64}{103} = \frac{39}{103},$$

$$P(100 \leq X < 130) = F(130) - F(100) = F(135) - F(105) = \frac{90}{103} - \frac{44}{103} = \frac{46}{103},$$

$$P(80 > X) = P(X < 80) = F(85) = \frac{17}{103}.$$

Obliczając powyższe prawdopodobieństwa na podstawie rozkładu zmiennej losowej X , uzyskujemy takie same wartości, co potwierdza poprawność wyników.

Chcąc wyznaczyć odchylenie standardowe, obliczamy najpierw wariancję z własności (5) wzoru (I.2.5):

$$E(X) = \sum_i x_i p_i = \frac{55 \cdot 4 + 65 \cdot 6 + 75 \cdot 7 + 85 \cdot 10 + \dots + 145 \cdot 6}{103} = \frac{10\,615}{103},$$

$$E(X^2) = \sum_i x_i^2 p_i = \frac{55^2 \cdot 4 + 65^2 \cdot 6 + 75^2 \cdot 7 + \dots + 135^2 \cdot 7 + 145^2 \cdot 6}{103} = \frac{1\,146\,975}{103},$$

$$V(X) = E(X^2) - E^2(X) = \frac{1\,146\,975 \cdot 103 - 10615^2}{103^2} = \frac{5\,460\,200}{103^2} \approx 514,7 \text{ (kWh)}^2.$$

Stąd odchylenie standardowe w próbie losowej średnich miesięcznych wartości zużycia energii wynosi:

$$\sigma(X) = \sqrt{V(X)} = 22,7 \text{ kWh.} \quad \blacksquare$$

II.2.2. Wartość średnia i miary rozproszenia

Jeżeli x_1, x_2, \dots, x_n jest n -elementową próbą statystyczną, to najpopularniejszym przybliżeniem wartości średniej w tej próbie jest **średnia arytmetyczna**:

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{II.2.3})$$

W przypadku gdy wyniki w próbie są uporządkowane w klasy i reprezentują je ich środki z uwzględnieniem liczności poszczególnych klas, wówczas średnia arytmetyczna jest zbyt słabym przybliżeniem wartości przeciętnej (średniej). Odpowiedniejsza w tym wypadku jest **średnia ważona**, uwzględniająca różne częstości występowania poszczególnych wartości w próbie losowej:

$$\hat{x} = \frac{\sum_{i=1}^n n_i \cdot x_i}{\sum_{i=1}^n n_i} = \frac{1}{n} \sum_{i=1}^n n_i \cdot x_i \quad (\text{II.2.4})$$

We wzorze (II.2.4) liczności n_i pełnią rolę tzw. **wag**.

Wzór (II.2.4) odpowiada definicji I.2.5 wartości przeciętnej, gdy prawdopodobieństwo jest określone według klasycznej definicji prawdopodobieństwa zdarzenia A , mówiącej, że jego wartość to iloraz liczby zdarzeń elementarnych sprzyjających zdarzeniu A i liczby wszystkich zdarzeń elementarnych w przestrzeni Ω . W przypadku szeregu rozdzielczego oznacza to, że:

$$p_i = \frac{n_i}{n} \quad (\text{II.2.5})$$

Drugim, obok wartości średniej, najważniejszym parametrem charakteryzującym próbę losową jest rozproszenie wyników w próbie. Najprostszą miarą rozproszenia jest, zdefiniowany w podrozdziale II.2.1, rozstęp. Jednakże nie informuje on o sposobie rozmieszczenia wartości w próbie, o ewentualnych tendencjach skupiania się wokół wartości średniej lub na końcach przedziału, gdyż uwzględnia tylko krańcowe wartości. Wszystkie elementy próby są natomiast uwzględnione w takich miarach rozproszenia jak wariancja $V(X)$ czy odchylenie standardowe $\sigma(X)$, zdefiniowanych w podrozdziale I.2.3. Odpowiednikami tamtych definicji dla próby statystycznej x_1, x_2, \dots, x_n jest:

$$V(X) = \sigma^2(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 \quad (\text{II.2.6})$$

a przy uwzględnieniu wag:

$$V(X) = \sigma^2(X) = \sum_{i=1}^n (x_i - \hat{x})^2 \cdot p_i \quad (\text{II.2.7})$$

Rolę odchylenia standardowego jako miary rozproszenia wyjaśnia przykład II.2.2.

PRZYKŁAD II.2.2

Handlowiec planuje hurtowy zakup pewnego produktu do swojego sklepu w ilości 100 kg. Produkt jest pakowany po 200 gramów. Dwóch producentów oferuje poszukiwany towar. U każdego z nich 500 opakowań produktu waży w sumie 100 kg, czyli średnia waga jednego opakowania równa jest nominalnej wartości 20 dkg. Jednakże odchylenia standardowe w obu partiach towaru okazały się różne: 20 gramów i 27 gramów. Można zatem przypuszczać, że pierwszy z producentów wykazuje większą precyzję w procesie produkcyjnym, gdyż odchylenie masy pojedynczego opakowania od nominalnych 200 gramów jest średnio mniejsze niż u drugiego producenta. Zatem kupujący składając zamówienie u pierwszego z producentów, ma większą szansę uniknąć ewentualnych przyszłych reklamacji ze strony swoich klientów. ■

Przy założeniu, że każdy element x_i z próby x_1, x_2, \dots, x_n może wystąpić z jednakowym prawdopodobieństwem $\frac{1}{n}$, wzory na momenty zwykłe α_r i centralne μ_r rzędu r , przyjmują analogiczną postać do wzorów z definicji I.2.6.:

$$\alpha_r = \frac{1}{n} \sum_{i=1}^n x_i^r, \quad \mu_r = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^r \quad (\text{II.2.8})$$

Definicja II.2.3. Zastępując potęgowaną zmienną jej wartością bezwzględną, zamiast momentów zwykłych i centralnych rzędu r , uzyskujemy tzw. **momenty zwykłe absolutne** i **momenty centralne absolutne rzędu r** :

$$a_r = \frac{1}{n} \sum_{i=1}^n |x_i|^r, \quad b_r = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}|^r \quad (\text{II.2.9})$$

Moment absolutny centralny pierwszego rzędu b_1 nazywa się **odchyleniem przeciętnym** od wartości średniej \hat{x} . Przy uwzględnieniu wag wzór przybiera postać następującą:

$$b_1 = \sum_{i=1}^n |x_i - \hat{x}| \cdot p_i \quad (\text{II.2.10})$$

Warto jeszcze wspomnieć o dwóch innych charakterystykach wyznaczanych z odchylenia standardowego σ oraz z odchylenia przeciętnego od średniej b_1 .

Są to:

$$1) \text{ zmienność } \lambda = \frac{\sigma(X)}{\hat{X}} \quad (\text{II.2.11})$$

$$2) \text{ nierównomierność } H = \frac{b_1}{\hat{X}} \quad (\text{II.2.12})$$

Przykład II.2.3 pokazuje stopień wrażliwości obu współczynników na różnice w rozproszeniu wartości w dwóch próbach statystycznych.

PRZYKŁAD II.2.3

Dane są dwie próby losowe:

X : 70, 35, 35, 70, 35, 70.

Y : 35, 70, 105, 70, 105, 35.

Porównajmy dla nich współczynniki zmienności i nierównomierności.

$$\hat{x} = 52,5$$

$$\hat{y} = 70,0,$$

$$\sigma(x) = 17,5,$$

$$\sigma(y) \approx 28,6,$$

$$b_1(x) = 17,5$$

$$b_1(y) \approx 23,3,$$

$$\lambda(x) \approx 0,333,$$

$$\lambda(y) \approx 0,408,$$

$$H(x) \approx 0,333,$$

$$H(y) \approx 0,333.$$

Na podstawie odchyłeń standardowych σ oraz odchyłeń przeciętnych b_1 widać, że wartości próby dla zmiennej Y są bardziej rozproszone. Jednakże wrażliwy na to okazał się tylko współczynnik zmienności λ . Nasuwa się zatem następujący wniosek: jeżeli w ramach kontroli pobieramy co pewien czas do badania n -elementowe próby pewnego produktu, współczynnik nierównomierności może nie wykryć rozregulowania się procesu technologicznego, objawiającego się w zwiększonym rozrzucie wartości w próbach. ■

II.3. Wybrane rozkłady z próby

W omawianych w następnej kolejności zagadnieniach estymacji i weryfikacji hipotez statystycznych (podrozdz. II.4 i II.5) wykorzystuje się rozkłady prawdopodobieństwa wybranych funkcji zmiennych losowych. Funkcje te nazywamy **statystykami**.

Definicja II.3.1. Statystyka to każda funkcja $g(X_1, X_2, \dots, X_n)$ zmiennych losowych X_1, X_2, \dots, X_n . Jako funkcja zmiennych losowych, statystyka jest również zmienną losową, a jej rozkład zależy od postaci funkcji g i od rozkładów zmiennych X_1, X_2, \dots, X_n .

Zwróćmy uwagę na trzy różne znaczenia terminu „rozkład”:

- 1) Rozkład cechy w populacji, zdefiniowany za pomocą parametrów, np.: rozkład normalny o wartości przeciętnej μ i odchyleniu standardowym σ : $N(\mu, \sigma)$, czy rozkład Studenta o k stopniach swobody T_k .
- 2) Rozkład wyników w próbie losowej (rozkład z próby), czyli rozkład otrzymany empirycznie. Gdy liczebność próby przekracza 30 elementów, $n > 30$, rozkład ten można uznać za reprezentatywny dla populacji.
- 3) Rozkład statystyki wyznaczonej na podstawie próby losowej, np.: rozkład średniej arytmetycznej z próby (twierdzenie II.3.1).

II.3.1. Rozkład średniej arytmetycznej z próby

Twierdzenie II.3.1

ZAŁOŻENIE

X_1, X_2, \dots, X_n – ciąg niezależnych zmiennych losowych o rozkładzie normalnym $N(\mu, \sigma)$ z parametrami μ, σ .

TEZA

Estymator \hat{X} wartości średniej, w postaci średniej arytmetycznej $\hat{x} = \frac{1}{n} \sum_{i=1}^n X_i$ ma rozkład normalny o parametrach $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Dowód

Ogólny kształt rozkładu zostaje zachowany ze względu na liniowy charakter przekształcenia X_i w \hat{X} .

Należy udowodnić, że:

- 1) wartość oczekiwana dla zmiennej \hat{X} wynosi μ ,
- 2) wariancja wynosi $\frac{\sigma^2}{n}$.

Ad 1)

$$E(\hat{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) \stackrel{E(X_i)=\mu}{=} \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \mu = \mu.$$

Ad 2)

$$\begin{aligned} V(\hat{X}) &= E\left[(\hat{X} - E(\hat{X}))^2\right] = E\left[(\hat{X} - \mu)^2\right] = E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{n}{n} \mu\right)^2\right] = \\ &= \frac{1}{n^2} E\left[\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right] = \frac{1}{n^2} E\left[\sum_{i=1}^n (X_i - \mu)^2 + 2 \sum_{\substack{i,j=1 \\ j \neq i}}^n (X_i - \mu)(X_j - \mu)\right] = \dots \end{aligned}$$

z liniowości wartości przeciętnej (własności I.2.4):

$$\dots = \frac{1}{n^2} \sum_{i=1}^n E[(X_i - \mu)^2] + \frac{2}{n^2} \sum_{\substack{i,j=1 \\ j \neq i}}^n E[(X_i - \mu)(X_j - \mu)] = \dots$$

z definicji wariancji i kowariancji (wzór I.3.7):

$$\dots = \frac{1}{n^2} \sum_{i=1}^n V(X_i) + \frac{2}{n^2} \sum_{\substack{i,j=1 \\ j \neq i}}^n \text{cov}(X_i, X_j) = \dots$$

z założenia o niezależności zmiennych losowych X_i, X_j dla $i \neq j$:

$$\dots = \frac{V(X_i) = \sigma^2}{n^2} \cdot n \sigma^2 + \frac{2}{n^2} \sum_{\substack{i,j=1 \\ j \neq i}}^n 0 = \frac{\sigma^2}{n}$$

Wykazaliśmy, że: $V(\hat{X}) = \frac{\sigma^2}{n}$. Z tego wynika: $\sigma(\hat{X}) = \frac{\sigma}{\sqrt{n}}$.

Uzyskujemy więc tezę twierdzenia: $\hat{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. c.b.d.u.

Zachodzi również następujące twierdzenie, wynikające z centralnego twierdzenia granicznego, które można znaleźć, wraz z dowodem, w pozycji [4]. Jest ono bardzo ważne z punktu widzenia praktycznych zastosowań:

Twierdzenie II.3.2 (Lindeberga–Levy’ego)

ZAŁOŻENIE

Zmienna X ma w populacji dowolny rozkład, zdefiniowany za pomocą wartości przeciętnej μ oraz wariancji σ^2 . Z populacji tej pobieramy próby losowe x_1, x_2, \dots, x_n o liczności n .

TEZA

Wraz ze wzrostem n , rozkład średniej z próby \hat{x} dąży do rozkładu normalnego

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Uwaga. W praktycznych zastosowaniach normalność rozkładu \hat{x} przyjmuje się dla $n > 30$.

II.3.2. Rozkład wariancji z próby

W podrozdziale I.2.5 podano sposób definiowania rozkładu χ^2 o k stopniach swobody, poprzez sumę kwadratów k zmiennych losowych Z_1, Z_2, \dots, Z_k o standaryzowanym rozkładzie normalnym. Wiemy też, że wartość oczekiwana dla rozkładu χ_k^2 wynosi k , a wariancja $2k$.

W podrozdziale II.2.2 wzór (II.2.6) definiuje jedną z możliwych statystyk dla oszacowania wariancji w próbie losowej x_1, x_2, \dots, x_n :

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2 \stackrel{\text{ozn.}}{=} \bar{S}^2 \quad (\text{II.3.1})$$

Jeżeli znana jest wartość przeciętna w próbie i nie musimy jej przybliżać liczeniem średniej \hat{x} , wówczas wzór na określenie wariancji przyjmuje postać:

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \stackrel{\text{ozn.}}{=} S^2 \quad (\text{II.3.2})$$

Często korzysta się również z innego określenia wariancji, ze względu na jego korzystne własności, które omówimy w podrozdziale II.4.1:

$$V(X) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2 \stackrel{\text{ozn.}}{=} \hat{S}^2 \quad (\text{II.3.3})$$

Dla statystyki \hat{S}^2 zachodzi następujące twierdzenie:

Twierdzenie II.3.3

ZAŁOŻENIE

X_1, X_2, \dots, X_n – ciąg niezależnych zmiennych losowych o rozkładzie $N(\mu, \sigma)$.

TEZA

Statystyka $\chi^2 = \frac{n-1}{\sigma^2} \cdot \hat{S}^2$ ma rozkład χ^2 o $(n-1)$ stopniach swobody.

Dowód

Na potrzeby dowodu przekształcamy statystykę χ^2 :

$$\chi^2 = \frac{n-1}{\sigma^2} \cdot \hat{S}^2 = \frac{n-1}{\sigma^2} \cdot \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2 = \sum_{i=1}^n \left(\frac{x_i - \hat{x}}{\sigma} \right)^2.$$

Uzyskaliśmy sumę kwadratów zmiennych losowych standaryzowanych, a zatem mających rozkład $N(0,1)$, czyli uzyskaliśmy zmienną o rozkładzie χ^2 . Ponieważ do obliczenia \hat{S}^2 potrzebna jest wartość \hat{x} , czyli dodatkowe równanie zawierające zaobserwowane wartości zmiennych losowych x_1, x_2, \dots, x_n , a zatem tracimy jeden stopień swobody. Czyli zmienna χ^2 ma $n-1$ stopni swobody.

c.b.d.u.

W sposób analogiczny dowodzi się następujące twierdzenia:

Twierdzenie II.3.4

ZAŁOŻENIE

X_1, X_2, \dots, X_n – ciąg niezależnych zmiennych losowych o rozkładzie $N(\mu, \sigma)$,

TEZA

Statystyka $\chi^2 = \frac{n}{\sigma^2} \cdot S^2$ ma rozkład χ^2 o n stopniach swobody.

Twierdzenie II.3.5

ZAŁOŻENIE

X_1, X_2, \dots, X_n – ciąg niezależnych zmiennych losowych o rozkładzie $N(\mu, \sigma)$,

TEZA

Statystyka $\chi^2 = \frac{n}{\sigma^2} \cdot \bar{S}^2$ ma rozkład χ^2 o $(n-1)$ stopniach swobody.

Zgodnie z definicją II.3.1, statystyki \hat{S}^2 , S^2 , \bar{S}^2 są również zmiennymi losowymi. Zatem można obliczyć ich wartości oczekiwane i wariancje. Z powyższych twierdzeń oraz z własności rozkładu χ_n^2 o n stopniach swobody, opisanych w podrozdziale I.2.5 wynika, że:

$$E\left(\frac{n-1}{\sigma^2} \hat{S}^2\right) = n-1, \quad V\left(\frac{n-1}{\sigma^2} \hat{S}^2\right) = 2(n-1) \quad (\text{II.3.4})$$

$$E\left(\frac{n}{\sigma^2} S^2\right) = n, \quad V\left(\frac{n}{\sigma^2} S^2\right) = 2n \quad (\text{II.3.5})$$

$$E\left(\frac{n}{\sigma^2} \bar{S}^2\right) = n-1, \quad V\left(\frac{n}{\sigma^2} \bar{S}^2\right) = 2(n-1) \quad (\text{II.3.6})$$

A zatem, wykorzystując własności wartości oczekiwanej (I.2.4) i wariancji (I.2.5), opisane w podrozdziale I.2.3, możemy wyznaczyć te dwa parametry dla trzech zdefiniowanych statystyk:

$$E(\hat{S}^2) = E\left(\frac{\sigma^2(n-1)}{\sigma^2(n-1)} \hat{S}^2\right) = \frac{\sigma^2}{n-1} E\left(\frac{n-1}{\sigma^2} \hat{S}^2\right) = \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2 \quad (\text{II.3.7a})$$

$$V(\hat{S}^2) = V\left(\frac{\sigma^2(n-1)}{\sigma^2(n-1)} \hat{S}^2\right) = \frac{\sigma^4}{(n-1)^2} V\left(\frac{n-1}{\sigma^2} \hat{S}^2\right) = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = \frac{2}{n-1} \sigma^4 \quad (\text{II.3.7b})$$

$$E(S^2) = E\left(\frac{n \sigma^2}{n \sigma^2} S^2\right) = \frac{\sigma^2}{n} E\left(\frac{n}{\sigma^2} S^2\right) = \frac{\sigma^2}{n} \cdot n = \sigma^2 \quad (\text{II.3.8a})$$

$$V(S^2) = V\left(\frac{n \sigma^2}{n \sigma^2} S^2\right) = \frac{\sigma^4}{n^2} V\left(\frac{n}{\sigma^2} S^2\right) = \frac{\sigma^4}{n^2} \cdot 2n = \frac{2}{n} \sigma^4 \quad (\text{II.3.8b})$$

$$E(\bar{S}^2) = E\left(\frac{n\sigma^2}{n\sigma^2} \bar{S}^2\right) = \frac{\sigma^2}{n} E\left(\frac{n}{\sigma^2} \bar{S}^2\right) = \frac{\sigma^2}{n} \cdot (n-1) = \frac{n-1}{n} \sigma^2 \quad (\text{II.3.9a})$$

$$V(\bar{S}^2) = V\left(\frac{n\sigma^2}{n\sigma^2} \bar{S}^2\right) = \frac{\sigma^4}{n^2} V\left(\frac{n}{\sigma^2} \bar{S}^2\right) = \frac{\sigma^4}{n^2} \cdot 2(n-1) = \frac{2(n-1)}{n^2} \sigma^4 \quad (\text{II.3.9b})$$

Wyniki powyższych obliczeń zamieszczono w tabeli II.3.1, przy założeniu $X \sim N(\mu, \sigma)$.

Tabela II.3.1

Wartości przeciętne i wariancje wybranych statystyk z próby losowej x_1, x_2, \dots, x_n

Statystyka	Wartość przeciętna	Wariancja
\hat{X}	μ	$\frac{\sigma^2}{n}$
\hat{S}^2	σ^2	$\frac{2}{n-1} \sigma^4$
S^2	σ^2	$\frac{2}{n} \sigma^4$
\bar{S}^2	$\frac{n-1}{n} \sigma^2$	$\frac{2(n-1)}{n^2} \sigma^4$

II.3.3. Rozkład ilorazu wariancji z prób

W podrozdziale I.2.5, zawierającym przykłady rozkładów ciągłych, jako ostatni został zdefiniowany rozkład F-Snedecora o k_1, k_2 stopniach swobody, jako rozkład ilorazu niezależnych zmiennych losowych X i Y o rozkładach χ^2 z – odpowiednio – k_1 i k_2 stopniami swobody

$$F = \frac{X/k_1}{Y/k_2} \quad (\text{II.3.10})$$

Uwaga: Tablice tego rozkładu są skonstruowane dla $F \geq 1$, zatem podczas praktycznych zastosowań, do licznika w definicji statystyki F zawsze wstawiamy zmienną o większej wartości.

Twierdzenie II.3.6

ZAŁOŻENIE

\hat{S}_x^2, \hat{S}_y^2 – wariancje z prób losowych x_1, x_2, \dots, x_n oraz y_1, y_2, \dots, y_m dla niezależnych zmiennych losowych X, Y , o rozkładach normalnych, dla których stosunek wariancji wynosi:

$$\frac{\sigma_y^2}{\sigma_x^2} = \lambda^2.$$

Statystyka $F = \lambda^2 \cdot \frac{\hat{S}_x^2}{\hat{S}_y^2}$ ma rozkład F-Snedecora o $(n-1, m-1)$ stopniach swobody.

Dowód

Na potrzeby dowodu przekształcamy statystykę F , rozpatrując dwa następujące warianty.

- 1) Zmienne X i Y mają tę samą wariancję σ^2 , czyli $\lambda = 1$:

$$F = \frac{\hat{S}_x^2}{\hat{S}_y^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2}{\frac{1}{m-1} \sum_{i=1}^m (y_i - \hat{y})^2} \cdot \frac{\sigma^2}{\sigma^2} = \frac{m-1}{n-1} \cdot \frac{\sum_{i=1}^n (x_i - \hat{x})^2}{\sum_{i=1}^m (y_i - \hat{y})^2} \cdot \frac{\sigma^2}{\sigma^2} =$$

$$= \frac{m-1}{n-1} \cdot \frac{n\bar{S}_x^2}{\sigma^2} \cdot \frac{\sigma^2}{m\bar{S}_y^2} = \frac{m-1}{n-1} \cdot \left[\frac{n\bar{S}_x^2}{\sigma^2} : \frac{m\bar{S}_y^2}{\sigma^2} \right] = \frac{\frac{n\bar{S}_x^2}{\sigma^2}}{\frac{m\bar{S}_y^2}{\sigma^2}} \cdot \frac{n-1}{m-1}$$

Zgodnie z twierdzeniem II.3.5, wyrażenia w nawiasie kwadratowym mają rozkłady χ^2 o $n-1$ oraz o $m-1$ stopniach swobody. Zatem ostateczny iloraz spełnia warunek (II.3.10) definicji rozkładu F-Snedecora o $(n-1, m-1)$ stopniach swobody.

- 2) Zmienne X i Y mają różne wariancje σ_x^2, σ_y^2 .

Z założenia mamy $\sigma_y^2 = \lambda^2 \cdot \sigma_x^2$:

$$F = \lambda^2 \cdot \frac{\hat{S}_x^2}{\hat{S}_y^2} = \lambda^2 \cdot \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2}{\frac{1}{m-1} \sum_{i=1}^m (y_i - \hat{y})^2} \cdot \frac{\sigma_y^2}{\lambda^2 \sigma_x^2} = \frac{m-1}{n-1} \cdot \frac{\sum_{i=1}^n (x_i - \hat{x})^2}{\sum_{i=1}^m (y_i - \hat{y})^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} =$$

$$= \frac{m-1}{n-1} \cdot \frac{n\bar{S}_x^2}{\sigma_x^2} \cdot \frac{\sigma_y^2}{m\bar{S}_y^2} = \frac{m-1}{n-1} \cdot \left[\frac{n\bar{S}_x^2}{\sigma_x^2} : \frac{m\bar{S}_y^2}{\sigma_y^2} \right] = \frac{\frac{n\bar{S}_x^2}{\sigma_x^2}}{\frac{m\bar{S}_y^2}{\sigma_y^2}} \cdot \frac{n-1}{m-1}$$

Uzyskany wynik jest taki sam jak w przypadku jednakowych wariancji. Zatem w tym przypadku, rozkład ilorazu również jest rozkładem F-Snedecora o $(n-1, m-1)$ stopniach swobody.

c.b.d.u.

II.3.4. Rozkład ilorazu wartości przeciętnej i odchylenia standardowego z próby

W podrozdziale I.2.5, jako jeden z przykładów rozkładów ciągłych, został zdefiniowany rozkład T-Studenta o k stopniach swobody. Jest on utworzony ze zmiennej o standaryzowanym rozkładzie normalnym $Z \sim N(0,1)$ oraz ze zmiennej o rozkładzie χ_k^2 dla k stopni swobody w następujący sposób:

$$T_k = \frac{Z}{\sqrt{\frac{\chi_k^2}{k}}} = \frac{Z}{\sqrt{\chi_k^2}} \sqrt{k} \quad (\text{II.3.11})$$

Udowodnimy teraz następujące twierdzenie:

Twierdzenie II.3.7

ZAŁOŻENIE

X_1, X_2, \dots, X_n – ciąg zmiennych losowych o rozkładzie normalnym $N(\mu, \sigma)$, \hat{X} oraz \bar{S}^2 są zdefiniowane wzorami (II.2.3) i (II.3.1).

TEZA

Statystyka $T = \frac{\hat{X} - \mu}{\bar{S}} \cdot \sqrt{n-1}$ ma rozkład T-Studenta o $(n-1)$ stopniach swobody.

Dowód

Na potrzeby dowodu przekształcamy statystykę T :

$$T = \frac{\hat{X} - \mu}{\bar{S}} \cdot \sqrt{n-1} = \left[\frac{\frac{\hat{X} - \mu}{\sigma}}{\frac{\bar{S}}{\sigma} \sqrt{n}} \sqrt{n} \right] \cdot \sqrt{n-1} = \frac{\frac{\hat{X} - \mu}{\sigma} \sqrt{n}}{\sqrt{\frac{\bar{S}^2 n}{\sigma^2}}} \cdot \sqrt{n-1}.$$

Z założenia i z twierdzeń II.3.1 oraz I.2.1 mamy, że $\frac{\hat{X} - \mu}{\sigma / \sqrt{n}}$ ma standaryzowany rozkład normalny $N(0, 1)$. Natomiast z twierdzenia II.3.5 wynika, że $\frac{\bar{S}^2 n}{\sigma^2}$ ma rozkład χ^2 o $(n-1)$ stopniach swobody. Zatem całe powyższe wyrażenie T spełnia warunki (II.3.11) definicji rozkładu T-Studenta dla $(n-1)$ stopni swobody.

c.b.d.u.

Wniosek. W analogiczny sposób można udowodnić, że statystyka $T = \frac{\hat{X} - \mu}{\hat{S}} \cdot \sqrt{n}$ ma rozkład Studenta o $(n-1)$ stopniach swobody dla \hat{S}^2 określonego wzorem (II.3.3).

Wystarczy zauważyć następującą tożsamość:

$$\frac{\hat{X} - \mu}{\bar{S}} \cdot \sqrt{n-1} = \frac{\hat{X} - \mu}{\hat{S}} \cdot \sqrt{n} \quad (\text{II.3.12})$$

Reasumując – statystykę T , mającą rozkład T-Studenta o $(n-1)$ stopniach swobody, można zapisać w następującej postaci:

$$T = \frac{\hat{X} - \mu}{\sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X})^2}{n(n-1)}}} \quad (\text{II.3.13})$$

II.3.5. Rozkład frakcji w próbie

Frakcja, zwana również **wskaźnikiem struktury**, jest prawdopodobieństwem wylosowania z populacji elementu mającego określoną własność. Na przykład pytamy jaki odsetek zakładów przemysłowych na określonym terenie nie przestrzega norm emisji do atmosfery substancji szkodliwych.

W podrozdziale I.2.4, jako jeden z przykładów rozkładów dyskretnych, został zdefiniowany rozkład zero-jedynkowy (dwupunktowy) z parametrem p . Parametr p w tym rozkładzie pełni rolę frakcji.

Na podstawie znajomości głównych charakterystyk w rozkładzie dwupunktowym $E(X) = p$, $V(X) = pq$, gdzie $q = 1 - p$, oraz na podstawie twierdzenia II.3.2, można udowodnić następujące twierdzenie:

Twierdzenie II.3.8

ZAŁOŻENIE

x_1, x_2, \dots, x_n – próba losowa zmiennej o rozkładzie dwupunktowym z parametrem p , k – liczba elementów w próbie losowej, posiadających wyróżnioną cechę.

TEZA

Statystyka $\hat{p} = \frac{k}{n}$ ma w przybliżeniu rozkład normalny $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Dowód pozostawiam Czytelnikowi, jako ćwiczenie.

II.4. Estymacja

Dwie najważniejsze metody wnioskowania statystycznego to:

- 1) estymacja,
- 2) weryfikacja (testowanie) hipotez statystycznych.

Estymacja to szacowanie nieznanych wartości parametrów rozkładu wyróżnionej cechy lub funkcji tych parametrów, na podstawie wybranych elementów badanej zbiorowości. Te wybrane, a właściwie wylosowane elementy to tzw. **próba statystyczna**. Zakłada się, że losowanie odbywa się zgodnie z rozkładem równomiernym, co pociąga za sobą przypadkowość uzyskiwanych wartości badanej cechy. A zatem możemy n -elementową próbę losową potraktować jako n -wymiarową zmienną losową (X_1, X_2, \dots, X_n) , w której poszczególne zmienne są niezależne i mają ten sam rozkład, zgodny z rozkładem badanej cechy w populacji generalnej. Naturalne jest, że wiarygodność wniosków dotyczących populacji, a wyciąganych na podstawie próbki, wzrasta wraz ze wzrostem liczebności próbki. Innymi słowy, wraz ze wzrostem liczebności próby rośnie dokładność (wiarygodność) przybliżenia nieznanego parametru przez estymator. W tym miejscu niezbędne jest wyjaśnienie pojęcia estymator.

Definicja II.4.1. Estymator parametru θ , to każda statystyka (funkcja wartości elementów próby losowej) $\hat{\theta} = \hat{\theta}_n(X_1, X_2, \dots, X_n)$, której wartość przyjmujemy jako oszacowanie nieznanego parametru θ .

II.4.1. Własności estymatorów

Dla jednego parametru można utworzyć wiele estymatorów. Wybór najlepszego wiąże się ze sprawdzeniem, który z nich posiada najlepsze własności. Oczywiście jest wymaganie, by wraz ze wzrostem liczebności próbki, rosła dokładność oszacowania parametru.

Do własności estymatorów należą:

- (1) zgodność – estymator nazywamy zgodnym, jeżeli spełniony jest następujący warunek:

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) = 1 \quad (\text{II.4.1})$$

który oznacza, że wraz ze wzrostem liczebności próby losowej, wzrasta dokładność oszacowania parametru θ ;

- (2) nieobciążoność – estymator jest nieobciążony, jeżeli jego wartość oczekiwana równa jest estymowanemu parametrowi:

$$\forall n \in N \quad E(\hat{\theta}_n) = \theta \quad (\text{II.4.2})$$

- (3) efektywność – **estymator efektywny** (najefektywniejszy) spośród wszystkich nieobciążonych estymatorów $\hat{\theta}_n$ parametru θ to ten, który ma najmniejszą wariancję, czyli o największym skupieniu jego wartości wokół szacowanego parametru.
- (4) dostateczność – **estymator dostateczny** $\hat{\theta}$ parametru θ skupia w sobie wszystkie informacje o parametrze θ zawarte w próbie losowej, czyli w praktyce jest wyznaczany ze wszystkich elementów próby.

Asymptotyczna nieobciążoność lub **asymptotyczna efektywność** estymatora oznacza odpowiednio nieobciążoność lub efektywność przy $n \rightarrow \infty$. Na przykład odchylenie standardowe $\sigma_n(X)$ (wzór (II.2.6)), obliczane jako pierwiastek kwadratowy z sumy kwadratów odchyleń od średniej podzielonej przez n (obciążona ocena wariancji), stanowi ocenę odchylenia standardowego w bieżącej próbie i nie można z niej wyciągać wniosków o populacji, z której próba została pobrana. Jednak jest to estymator asymptotycznie nieobciążony. Natomiast gdy sumę kwadratów dzielimy przez $n-1$, a nie n (nieobciążona ocena wariancji), to wynikająca stąd wartość $\sigma_{n-1}(X)$ stanowi opis wartości odchylenia standardowego w populacji. A zatem taka statystyka może być następnie używana do wyciągania wniosków o całej populacji, z której pobrano próbę. Można wykazać, że oba estymatory $\sigma_n(X)$, $\sigma_{n-1}(X)$ są estymatorami zgodnymi.

Jeżeli dla tego samego parametru θ , znajdziemy dwa estymatory zgodne i nieobciążone, wówczas za lepszy z nich uznajemy ten o mniejszej wariancji, czyli bardziej efektywny, gdyż dla niego mamy większe skupienie jego wartości wokół wartości nieznanego parametru θ .

II.4.2. Metody estymacji

W wyniku estymacji uzyskujemy tzw. estymatory, czyli przybliżone wartości szukanych parametrów (charakterystyk liczbowych).

II.4.2.1. Metoda najmniejszych kwadratów

Spośród kilku metod estymacyjnych zdecydowanie najczęściej znajduje zastosowanie klasyczna metoda najmniejszych kwadratów (MKN), m.in. dlatego, że otrzymywane estymatory są asymptotycznie zgodne z estymatorami otrzymywanymi metodą największej wiarygodności, omówioną w następnym podrozdziale (pod warunkiem, że wektor składnika losowego ma rozkład normalny), oraz są najlepszymi estymatorami wśród wszystkich estymatorów liniowych. Istnieje twierdzenie, znane jako twierdzenie Markowa, mówiące, że estymatory uzyskane tą metodą są zgodne, nieobciążone i najefektywniejsze. W twierdzeniu tym nie zakłada się normalności rozkładów zmiennych losowych, ani nawet ich niezależności.

Metoda najmniejszych kwadratów zmierza do minimalizacji sumy kwadratów odchyleń obserwowanych wartości zmiennej losowej x_i od jej wartości modelowych \bar{x}_i , czyli:

$$\sum_{i=1}^n (x_i - \bar{x}_i)^2 = \delta^T \delta = \min \quad (\text{II.4.3})$$

Niech X_1, X_2, \dots, X_n będzie ciągiem zmiennych losowych, a x_1, x_2, \dots, x_n – ciągiem ich wartości zaobserwowanych, czyli wartości z próby. Rozkład zmiennych losowych X_i zależy od nieznanymi parametrów $\theta_1, \theta_2, \dots, \theta_k$, których wartości trzeba oszacować (wyestymować).

Zakładamy również znajomość postaci zależności funkcyjnej zmiennych losowych X_i i parametrów θ_j :

$$X_i = g_i(\theta_1, \theta_2, \dots, \theta_k) \quad (\text{II.4.4})$$

Postać funkcji g zależy od specyfiki zagadnienia. Jeżeli funkcja g nie jest liniowa, to na potrzeby estymacji sprowadzamy ją do postaci liniowej za pomocą rozwinięcia w szereg Taylora.

Poniższe przykłady wyjaśniają zastosowanie MNK do wyznaczenia parametrów θ_i w konkretnych zagadnieniach.

PRZYKŁAD II.4.1

Dla przykładu wykonamy metodą najmniejszych kwadratów estymację parametrów równania regresji Y względem X :

$$y = ax + b \quad (\text{II.4.5})$$

przy założeniu, że (X, Y) jest dwuwymiarową zmienną losową o rozkładzie normalnym.

Rozwiązanie

Parametry a i b prostej (II.4.5) oraz parametry dwuwymiarowego rozkładu normalnego, są nieznane. Dla ich oszacowania zrealizowano n -elementową serię doświadczenia, w wyniku czego otrzymano n par liczbowych:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (\text{II.4.6})$$

stanowiących realizację zmiennej losowej (X, Y) . Możemy powiedzieć, że jest to próba losowa przy założeniu, że:

- prawdopodobieństwo znalezienia się w próbie było jednakowe dla każdego elementu populacji,
- wszystkie doświadczenia w serii były wykonane niezależnie.

Zgodnie z zasadą MNK, dla wyznaczenia parametrów a i b równania regresji, należy zminimalizować wyrażenie:

$$F = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n (y_i - (\hat{a}x_i + \hat{b}))^2 \quad (\text{II.4.7})$$

w którym \hat{a} i \hat{b} oznaczają estymatory parametrów a i b .

Warunkiem koniecznym minimalizacji funkcji jest zerowanie się pierwszych pochodnych cząstkowych po niewiadomych (zakładamy różniczkowalność funkcji F względem nieznanymi parametrów):

$$\begin{aligned} \frac{\partial F}{\partial \hat{a}} &= -2 \sum_{i=1}^n (y_i - \hat{a}x_i - \hat{b})x_i = 0 \\ \frac{\partial F}{\partial \hat{b}} &= -2 \sum_{i=1}^n (y_i - \hat{a}x_i - \hat{b}) = 0 \end{aligned} \quad (\text{II.4.8})$$

Równania (II.4.8) tworzą tzw. układ równań normalnych:

$$\begin{aligned} \sum_{i=1}^n x_i y_i - \hat{a} \sum_{i=1}^n x_i^2 - \hat{b} \sum_{i=1}^n x_i &= 0 \\ \sum_{i=1}^n y_i - \hat{a} \sum_{i=1}^n x_i - n \cdot \hat{b} &= 0 \end{aligned} \quad (\text{II.4.9})$$

Rozwiązując ten układ względem \hat{a} i \hat{b} , otrzymujemy:

$$\hat{b} = \hat{y} - \hat{a} \cdot \hat{x} \quad \text{oraz} \quad \hat{a} = \frac{\sum_{i=1}^n (x_i - \hat{x})(y_i - \hat{y})}{\sum_{i=1}^n (x_i - \hat{x})^2} \quad (\text{II.4.10})$$

W ten sposób, w wyniku zastosowania metody najmniejszych kwadratów, uzyskaliśmy estymatory parametrów linii regresji Y względem X z próby losowej. Z powyższych równań wynika, że linia regresji przechodzi przez punkt o współrzędnych (\hat{x}, \hat{y}) , zwany środkiem ciężkości. Analogicznie uzyskujemy parametry linii regresji X względem Y . ■

PRZYKŁAD II.4.2

Doświadczalne pole uprawy rolnej zostało podzielone na 60 prostokątów. Dla każdego z nich pomierzono i zapisano liczbę kilogramów zastosowanego nawozu sztucznego X i zebranego plonu Y (tab. II.4.1). Na podstawie wartości zmiennych X i Y dla wylosowanych 10 różnych prostokątów testowych, wyznaczyć linię regresji zależności plonu od ilości zastosowanego nawozu oraz zaprognozować plon, gdy użyje się 50 kg nawozu.

Tabela II.4.1

Wartości zmiennych losowych X, Y w próbie losowej 10-elementowej

X	18	20	25	36	40	55	61	87	90	91
Y	620	720	770	800	805	950	952	970	990	999

Rozwiązanie

Na podstawie treści zadania wiemy, że chodzi o wyznaczenie linii regresji Y względem X , czyli $y = ax + b$. W pierwszej kolejności wyznaczamy wartości średnie:

$$\hat{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 52,3 \quad \text{oraz} \quad \hat{y} = \frac{\sum_{i=1}^{10} y_i}{10} = 857,6.$$

Następnie wyznaczamy estymatory parametrów linii regresji a i b ze wzoru (II.4.10):

$$\hat{a} = \frac{32006,20}{7588,10} = 4,2179, \quad \hat{b} = 857,6 - 4,2179 \cdot 52,3 = 637,00.$$

Obliczenia zamieszczono w tabeli II.4.2.

Tabela II.4.2

Obliczenia pomocnicze przy wyznaczaniu parametrów linii regresji $Y(X)$

x_i	y_i	$x_i - \hat{x}$	$y_i - \hat{y}$	$(x_i - \hat{x}) \cdot (y_i - \hat{y})$	$(x_i - \hat{x})^2$
18	620	-34,3	-237,6	8149,68	1176,49
20	720	-32,3	-137,6	4444,48	1043,29
25	770	-27,3	-87,6	2391,48	745,29
36	800	-16,3	-57,6	938,88	265,69
40	805	-12,3	-52,6	646,98	151,29
55	950	2,7	92,4	249,48	7,29
61	952	8,7	94,4	821,28	75,69
87	970	34,7	112,4	3900,28	1204,09
90	990	37,7	132,4	4991,48	1421,29
91	999	38,7	141,4	5472,18	1497,69
Suma				32006,20	7588,10

Zatem linia regresji wielkości plonu względem stopnia nawożenia przyjmuje postać:

$$y = 4,2179 \cdot x + 637,00.$$

Na jej podstawie dokonujemy prognozy:

$$y(50) = 4,2179 \cdot 50 + 637,00 = 847,9,$$

czyli po zastosowaniu 50 kg nawozu powinniśmy otrzymać 848 kg plonu.

■

II.4.2.2. Metoda największej wiarygodności

Estymatory uzyskane metodą największej wiarygodności (MNW) również mają korzystne własności. Są to estymatory zgodne, asymptotycznie nieobciążone i asymptotycznie efektywne oraz mają rozkład asymptotycznie normalny.

Podobnie jak w metodzie MNK założmy, że X jest zmienną losową o rozkładzie zależnym od nieznanach parametrów $\theta_1, \theta_2, \dots, \theta_k$, których wartości trzeba oszacować. Zmienna X jest typu ciągłego o określonej funkcji gęstości f , lub typu skokowego o określonej funkcji rozkładu prawdopodobieństwa f . Niech x_1, x_2, \dots, x_n będzie ciągiem zaobserwowanych wartości zmiennej X .

Oznaczmy przez L tzw. **funkcję wiarygodności**, określoną wzorem:

$$L = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k) \quad (\text{II.4.11})$$

Metoda największej wiarygodności polega na tym, że estymatorami parametrów $\theta_1, \theta_2, \dots, \theta_k$, będą takie wartości $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$, dla których funkcja L osiąga maksimum. Wartości

$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ będą oczywiście zależeć od wartości zaobserwowanych w próbie losowej. Są więc ich funkcjami, czyli statystykami. Innymi słowy: poszukujemy takiego oszacowania nieznanych parametrów, dla którego prawdopodobieństwo otrzymania zaobserwowanych wartości jest największe.

Poszukując maksimum funkcji L , korzystamy z faktu, że funkcja $\ln L$ osiąga maksimum w tym samym punkcie co L . Podobnie jak dla funkcji F w MNK zakładamy różniczkowalność funkcji L względem nieznanych parametrów $\theta_1, \theta_2, \dots, \theta_k$. Zapisując warunek konieczny istnienia ekstremum, otrzymujemy układ k równań o k niewiadomych:

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \quad \text{dla } i = 1, 2, \dots, k \quad (\text{II.4.12})$$

Warunek wystarczający to ujemnie określona forma kwadratowa, utworzona z drugich pochodnych cząstkowych funkcji $\ln L$:

$$\sum_{i=1}^k \sum_{j=1}^k \left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right)_{\substack{\theta = \hat{\theta} \\ \theta_j = \hat{\theta}_j}} < 0 \quad (\text{II.4.13})$$

Estymatory uzyskiwane metodą największej wiarygodności mają dużą efektywność, ale mogą być obciążone, co pokaże poniższy przykład.

PRZYKŁAD II.4.3

Niech cecha X ma rozkład normalny $N(\mu, \sigma)$ o nieznanych parametrach μ i σ . Znajdziemy estymatory tych parametrów metodą MNW. Niech x_1, x_2, \dots, x_n będzie ciągiem zaobserwowanych wartości zmiennej X . Funkcja wiarygodności L ma w tym przypadku następującą postać:

$$L = \prod_{i=1}^n f(x_i; \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}} \quad (\text{II.4.14})$$

Logarytmujemy funkcję L :

$$\begin{aligned} \ln L &= \ln \left(\frac{e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}}}{(\sigma^2 \cdot 2\pi)^{n/2}} \right) = -\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} - \frac{n}{2} \cdot \ln(2\pi\sigma^2) = \\ &= -\frac{n}{2} \cdot \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned} \quad (\text{II.4.15})$$

Układ warunków (II.4.12) przyjmuje postać:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0, \quad \frac{\partial \ln L}{\partial \sigma} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma} = 0 \quad (\text{II.4.16})$$

Rozwiązując powyższy układ równań, uzyskujemy następujące estymatory parametrów μ i σ :

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \hat{x} \quad \text{oraz} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \hat{x})^2}{n} \stackrel{\text{ozn.}}{=} \bar{S}^2 \quad (\text{II.4.17})$$

Odpowiadają one określeniom (II.2.3) i (II.2.6) estymatorów wartości średniej i wariancji. Wykazaliśmy zatem, że średnia arytmetyczna wyników z próby i wariancja z próby są estymatorami największej wiarygodności nieznanego wartości μ i σ^2 rozkładu $N(\mu, \sigma)$. Można wykazać, że \hat{x} jest estymatorem zgodnym, nieobciążonym i efektywnym, a wariancja \bar{S}^2 jest estymatorem zgodnym i asymptotycznie nieobciążonym.

II.4.3. Estymacja punktowa

Estymacja punktowa to oszacowanie wartości nieznanego parametru θ , wraz z wyznaczeniem dokładności tego oszacowania. Zatem w wyniku estymacji punktowej uzyskujemy estymator i jego odchylenie standardowe: $\theta \approx \hat{\theta} \pm \sigma(\hat{\theta})$.

W badaniach statystycznych najczęściej poszukuje się estymatorów wartości przeciętnej i wariancji badanej cechy, gdy jest to cecha mierzalna, lub estymatorów frakcji, gdy jest to cecha niemierzalna. Frakcja jest to tzw. **względna częstość** pojawiania się pewnej cechy w populacji.

Na przykład w rozkładzie Bernoullego rolę frakcji pełni parametr p , czyli prawdopodobieństwo wylosowania w jednej próbie elementu mającego określoną cechę. W sytuacji, gdy szacujemy p z próby n -elementowej, to estymatorem zgodnym, nieobciążonym i efektywnym dla p jest częstość:

$$\hat{p} = \frac{k}{n} \quad (\text{II.4.18})$$

gdzie:

- k – liczba elementów w próbie losowej, mających określoną własność,
- n – liczność próby.

Taka postać estymatora wskaźnika struktury jest w pełni zgodna z intuicją, gdyż właśnie z częstością mamy do czynienia w klasycznej definicji prawdopodobieństwa.

Tabela II.4.3 zawiera wzory na podstawowe estymatory cech mierzalnych (ilościowych) oraz na ich odchylenia standardowe, wyznaczone na podstawie n -elementowej próby losowej x_1, x_2, \dots, x_n .

Tabela II.4.3
Estymatory i ich własności

Estymowany parametr	Estymator	Dokładność estymatora	Własności estymatora
Wartość przeciętna μ dla obserwacji jednakowo dokładnych (średnia arytmetyczna)	$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$\sigma(\hat{x}) = \frac{\sigma}{\sqrt{n}}$ ewentualnie:	zgodny, dostateczny, nieobciążony, dla rozkładu normalnego również efektywny
Wartość przeciętna μ dla obserwacji różnodokładnych (średnia ważona)	$\hat{x} = \frac{\sum_{i=1}^n p_i \cdot x_i}{\sum_{i=1}^n p_i}$	$\frac{S}{\sqrt{n-1}}$ lub $\frac{\bar{S}}{\sqrt{n-1}}$ lub $\frac{\hat{S}}{\sqrt{n}}$	
Wariancja σ^2 gdy μ znane	$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$	$\sigma(S^2) = \sqrt{\frac{2}{n}} \cdot S^2$	zgodny, dostateczny, nieobciążony, dla rozkładu normalnego również efektywny
Wariancja σ^2 gdy μ nieznanne dla obserwacji jednakowo dokładnych	$\bar{S}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2$	$\sigma(\bar{S}^2) = \frac{\sqrt{2 \cdot (n-1)}}{n} \cdot \bar{S}^2$	zgodny, dostateczny, asymptotycznie nieobciążony
	$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2$	$\sigma(\hat{S}^2) = \sqrt{\frac{2}{n-1}} \cdot \hat{S}^2$	zgodny, dostateczny, nieobciążony, dla rozkładu normalnego również asymptotycznie efektywny
	$\hat{S}^2 = \frac{1}{\sum_{i=1}^n p_i} \sum_{i=1}^n p_i (x_i - \hat{x})^2$		
Wariancja σ^2 gdy μ nieznanne dla obserwacji różnodokładnych			
Frakcja (wskaźnik struktury)	$\hat{p} = \frac{k}{n}$	$\sigma(\hat{p}) = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$	zgodny, dostateczny, nieobciążony, efektywny
Współczynnik zmienności	$\lambda = \frac{\sigma(X)}{\hat{X}}$	–	zgodny, dostateczny

II.4.4. Estymacja przedziałowa

Metody estymacji omówione powyżej w podrozdziale II.4.2 pozwalają uzyskiwać punktowe oceny nieznanego parametru rozkładu. Natomiast estymacja przedziałowa polega na znalezieniu takiego przedziału liczbowego (θ_1, θ_2) , dla nieznanego parametru θ , w którym ten parametr znajduje się z, zadany z góry, prawdopodobieństwem p . Przedział ten nazywa się **przedziałem ufności**, a prawdopodobieństwo p – **poziomem ufności**. Końce przedziału ufności są funkcjami próby losowej x_1, x_2, \dots, x_n i nie zależą od estymowanego parametru. A zatem są to zmienne losowe. Różnica $\theta_2 - \theta_1$ jest to tzw. **szerokość przedziału ufności**.

Oczywiście może się zdarzyć, że rzeczywista wartość parametru θ nie należy do wyznaczonego przedziału ufności. Jednakże wyznaczając przedziały ufności dla różnych prób losowych x_1, x_2, \dots, x_n , można przyjąć, że częstość tych, które będą zawierać rzeczywistą wartość θ , w przybliżeniu będzie równa p .

Konstrukcja przedziału ufności w określonym przypadku opiera się na twierdzeniach z rozdziału II.3, czyli na znajomości rozkładów określonych statystyk, zbudowanych z próby losowej. We wszystkich przytoczonych poniżej modelach tworzenia przedziałów ufności przyjmujemy oznaczenie liczności próby jako n oraz definiujemy $\alpha = 1 - p$.

II.4.4.1. Estymacja przedziałowa wartości przeciętnej

W zależności od tego, jakie warunki spełnia analizowana zmienna losowa oraz pobrana próba losowa, wyróżniamy trzy modele tworzenia symetrycznych przedziałów ufności dla wartości oczekiwanej.

Model I (na podstawie twierdzenia I.2.1)

ZAŁOŻENIA

$X \sim N(\mu, \sigma)$,

σ – znane,

n – dowolne.

Postać przedziału ufności:

$$\mu \in \left[\hat{X} - z \left(1 - \frac{\alpha}{2} \right) \cdot \sigma(\hat{X}); \hat{X} + z \left(1 - \frac{\alpha}{2} \right) \cdot \sigma(\hat{X}) \right] \quad (\text{II.4.19})$$

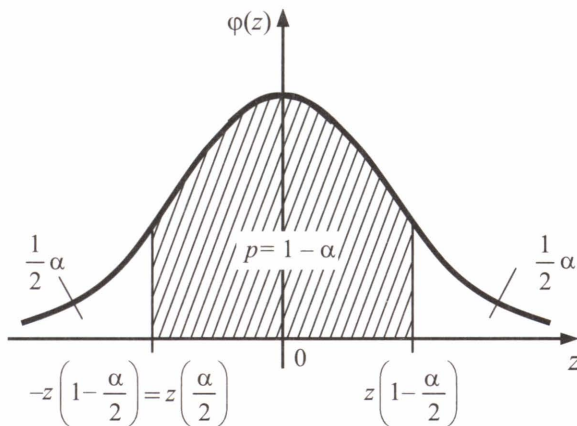
gdzie:

$$\alpha = 1 - p,$$

$z \left(1 - \frac{\alpha}{2} \right)$ – kwantyl standaryzowanego rozkładu normalnego,

$$\sigma(\hat{X}) = \frac{\sigma}{\sqrt{n}} \text{ (por. twierdzenie II.3.1).}$$

Szerokość powyższego przedziału ufności wynosi $2 \cdot z \left(1 - \frac{\alpha}{2} \right) \cdot \frac{\sigma}{\sqrt{n}}$ (rys. II.4.1). Zatem nie zależy ona od wartości z pobranej próby x_i , a jedynie od liczności próby n i poziomu ufności p .



Rys. II.4.1. Symetryczny przedział ufności dla wartości przeciętnej w próbie o znanym odchyleniu standardowym

Model II (na podstawie twierdzenia II.3.7)

ZAŁOŻENIA

$X \sim N(\mu, \sigma)$,

σ – nieznane,

n – dowolne (w szczególności małe $n < 30$).

Postać przedziału ufności:

$$\mu \in \left[\hat{X} - t\left(1 - \frac{\alpha}{2}, k\right) \cdot \sigma(\hat{X}); \hat{X} + t\left(1 - \frac{\alpha}{2}, k\right) \cdot \sigma(\hat{X}) \right] \quad (\text{II.4.20})$$

gdzie:

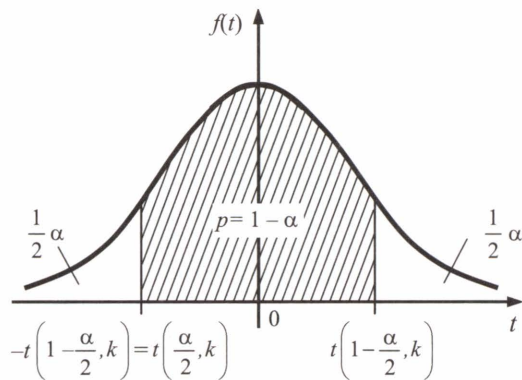
$$\alpha = 1 - p,$$

$t\left(1 - \frac{\alpha}{2}, k\right)$ – kwantyl rozkładu Studenta o k stopniach swobody,

$$\sigma(\hat{X}) = \frac{\hat{S}}{\sqrt{n}} \quad \text{lub} \quad \sigma(\hat{X}) = \frac{\bar{S}}{\sqrt{n-1}} \quad (\text{por. wzory (II.3.1) oraz (II.3.3)}).$$

W tym modelu szerokość przedziału ufności (rys. II.4.2) wynosi $2 \cdot t\left(1 - \frac{\alpha}{2}, k\right) \cdot \frac{\hat{\sigma}}{\sqrt{n}}$.

Zatem jest zmienną losową, bowiem jest estymowana na podstawie wartości z próby losowej.



Rys. II.4.2. Symetryczny przedział ufności dla wartości przeciętnej w mało licznej próbie o nieznanym odchyleniu standardowym

Model III (na podstawie twierdzenia II.3.2)

ZAŁOŻENIA

X – ma rozkład dowolny,

σ – nieznane, skończone,

n – duże ($n > 30$).

Postać przedziału ufności jak w modelu I, przy czym $\sigma(\hat{X}) = \frac{\hat{S}}{\sqrt{n}}$ lub $\sigma(\hat{X}) = \frac{\bar{S}}{\sqrt{n-1}}$.

PRZYKŁAD II.4.4

W tabeli II.4.4 zamieszczono średnie dobowe wyniki pomiaru stężenia pyłu zawieszonego w powietrzu, na terenie pewnego miasta przemysłowego. Wartości te podane są w $\mu\text{g}/\text{m}^3$. Pomiar był wykonywany codziennie w ciągu jednego miesiąca. Zmienna T oznacza kolejne dni w analizowanym miesiącu. Wykonać estymację punktową i przedziałową na poziomie ufności $p = 0,90$ dla średniego stężenia pyłu zawieszonego w analizowanym miesiącu, na badanym terenie.

Tabela II.4.4
Średnie dobowe stężenia pyłu zawieszonego

T	Pył	T	Pył	T	Pył	T	Pył	T	Pył
1	32,3	8	21,8	15	49,5	22	26,1	29	56,2
2	41,2	9	17,9	16	31,2	23	38,1	30	38,4
3	20,4	10	16,9	17	45,1	24	29,3	31	45,2
4	31,9	11	39,2	18	19,6	25	30,9		
5	24,2	12	25,9	19	33,5	26	29,8		
6	50,4	13	31,1	20	57,8	27	28,3		
7	28,8	14	55,7	21	42,1	28	48,6		

Rozwiązanie

Ponieważ nie ma żadnych przesłanek, by różnicować zaufanie do wyników zamieszczonych w tabeli II.4.4 (pomiarzy najprawdopodobniej zostały wykonane tym samym przyrządem, w tych samych warunkach), więc możemy je potraktować jako zbiór obserwacji jednokowo dokładnych. Zatem w celu wykonania estymacji punktowej korzystamy ze wzorów na estymator wartości przeciętnej (średniej) i nieobciążony estymator wariancji, zamieszczonych w tabeli II.4.3 dla takich właśnie obserwacji. Zmienna X reprezentuje średnie dobowe stężenie pyłu.

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{31} \sum_{i=1}^{31} x_i = \frac{1087,4}{31} = 35,077 \text{ } \mu\text{g}/\text{m}^3,$$

$$\hat{\sigma}^2 = \hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2 = \frac{1}{30} \sum_{i=1}^{31} (x_i - 35,077)^2 = \frac{4095,374}{30} = 136,5125 \text{ } \mu\text{g}^2/\text{m}^6.$$

Stąd odchylenie standardowe w próbie losowej złożonej z 31 obserwacji wynosi:

$$\hat{\sigma} = \sqrt{136,5125} = 11,684 \text{ } \mu\text{g}/\text{m}^3.$$

Powyższe odchylenie standardowe jest potrzebne do wyznaczenia dokładności estymatora wartości średniej, czyli $\sigma(\hat{x})$. Bowiem pełnym wynikiem estymacji punktowej jest podanie przybliżonej wartości estymowanego parametru wraz z oszacowaniem jego odchylenia standardowego, stanowiącego ocenę jego dokładności.

$$\sigma(x) = \frac{\hat{\sigma}}{\sqrt{n}} = \sqrt{\frac{136,5125}{31}} = 2,098 \text{ } \mu\text{g}/\text{m}^3.$$

Stąd ostateczny wynik estymacji punktowej średniego stężenia pyłu zawieszonego na badanym terenie, w analizowanym miesiącu jest następujący:

$$\mu \cong 35,1 \pm 2,1 \text{ } \mu\text{g}/\text{m}^3.$$

Dla wykonania estymacji przedziałowej posłużymy się modelem III ze względu na wystarczająco dużą licznosc próby losowej ($n > 30$), wykorzystując wzór (II.4.19) do wyliczenia granic przedziału ufności. Potrzebny kwantyl rozkładu normalnego $z\left(1 - \frac{\alpha}{2}\right) = z(0,95) = 1,6448$ odczytujemy z tablicy 2 (dodatek) dla $\alpha = 1 - p = 1 - 0,90 = 0,10$:

$$z\left(1 - \frac{\alpha}{2}\right) \cdot \sigma(\hat{X}) = 1,6448 \cdot 2,098 = 3,451,$$

$$\hat{X} - z\left(1 - \frac{\alpha}{2}\right) \cdot \sigma(\hat{X}) = 35,077 - 3,451 = 31,626,$$

$$\hat{X} + z\left(1 - \frac{\alpha}{2}\right) \cdot \sigma(\hat{X}) = 35,077 + 3,451 = 38,528.$$

Stąd wynik estymacji przedziałowej na poziomie ufności 90%, dla średniego stężenia pyłu zawieszonego w powietrzu na badanym terenie, w analizowanym miesiącu jest następujący:

$$\mu \in [31,6; 38,5] [\mu\text{g}/\text{m}^3].$$

Szerokość wyznaczonego przedziału ufności wynosi $6,9 \mu\text{g}/\text{m}^3$. Ostateczne wyniki estymacji punktowej i przedziałowej podajemy z dokładnością taką samą jak dokładność danych pomiarowych, czyli $0,1 \mu\text{g}/\text{m}^3$.

■

II.4.4.2. Estymacja przedziałowa wariancji i odchylenia standardowego

Przedziały ufności dla wariancji powstają podobnie jak dla wartości przeciętnej, tzn. z wykorzystaniem rozkładu wybranych statystyk. W tym wypadku modele są zróżnicowane tylko ze względu na licznosc próby. Zawsze zakładamy normalność rozkładu zmiennej losowej X oraz nieznanosc obu parametrów μ i σ :

Model I (na podstawie twierdzenia II.3.3)

ZAŁOŻENIA

$$X \sim N(\mu, \sigma),$$

μ, σ – nieznanne,

n – małe ($n < 50$).

Postać przedziału ufności dla wariancji:

$$\sigma^2 \in \left[\frac{k \cdot \hat{S}^2}{\chi^2 \left(1 - \frac{\alpha}{2}, k \right)}, \frac{k \cdot \hat{S}^2}{\chi^2 \left(\frac{\alpha}{2}, k \right)} \right] \quad (\text{II.4.21})$$

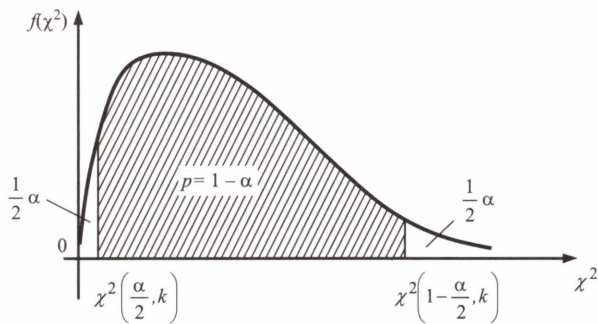
gdzie:

$$\alpha = 1 - p,$$

$\chi^2 \left(1 - \frac{\alpha}{2}, k \right), \chi^2 \left(\frac{\alpha}{2}, k \right)$ – kwantyle rozkładu chi-kwadrat o k stopniach swobody,

\hat{S}^2 – określone wzorem (II.3.3).

Ze względu na niesymetryczność rozkładu chi-kwadrat, przedział ufności na rysunku II.4.3 również nie jest symetryczny w swoim kształcie, ale zachowuje symetrię rozkładu pola powierzchni pod wykresem funkcji gęstości.



Rys. II.4.3. Symetryczny przedział ufności dla wariancji w próbie mało licznej

Model II (na podstawie twierdzenia II.3.2. Lindeberga–Levy’ego)

ZAŁOŻENIA

$$X \sim N(\mu, \sigma),$$

μ, σ – nieznane,

n – duże ($n > 50$).

Przedział ufności dla odchylenia standardowego:

$$\sigma \in \left[\frac{\hat{S}}{z\left(1 - \frac{\alpha}{2}\right) + \frac{1}{\sqrt{2 \cdot k}}}, \frac{\hat{S}}{z\left(1 - \frac{\alpha}{2}\right) - \frac{1}{\sqrt{2 \cdot k}}} \right] \quad (\text{II.4.22})$$

gdzie:

$$\alpha = 1 - p,$$

$z\left(1 - \frac{\alpha}{2}\right)$ – kwantyl standaryzowanego rozkładu normalnego,

\hat{S} – określone wzorem (II.3.3).

Z postaci przedziału (II.4.22) dla σ uzyskujemy przedział ufności dla wariancji, przez podniesienie końców przedziału do kwadratu.

PRZYKŁAD II.4.5

Wykonano wielokrotny pomiar ciśnienia atmosferycznego przy użyciu dwóch barometrów tej samej klasy dokładności. W obu przypadkach, w tych samych zewnętrznych warunkach pomiaru, wykonano 80 obserwacji.

Uzyskano następujące oszacowanie parametrów σ_1 , σ_2 , opisujących dokładności obu barometrów:

$$\hat{\sigma}_1 = 0,8 \text{ Pa}, \quad \hat{\sigma}_2 = 0,7 \text{ Pa}.$$

Wykorzystując estymację przedziałową, sprawdzić, czy na poziomie ufności 95% można uznać, że oba barometry dają faktycznie tę samą dokładność.

Rozwiązanie

Wiedząc, że estymacja przedziałowa ma na celu wyznaczenie przedziału dla nieznanego parametru, w którym ten parametr znajdzie się z zadaniem z góry prawdopodobieństwem i porównując przedziały wyznaczone dla dokładności pomiaru obu barometrów, możemy wnioskować nt. istotności różnic między nimi. I tak: jeśli wyznaczone przedziały są rozłączne, wówczas wiemy, że na zadanym poziomie ufności (0,95) oba przyrządy pomiarowe mierzą z różną dokładnością.

W przeciwnym wypadku uznajemy dokładności za zbliżone. Oczywiście przedziały mogą zachodzić na siebie w różnym stopniu, ale wykorzystując tylko wynik estymacji, nie jesteśmy w stanie ocenić wpływu wielkości obszaru nachodzenia na to, jaki błąd popełniamy, uznając parametry za zbliżone.

Ze względu na dużą liczebność próby losowej korzystamy z modelu II do wyznaczenia przedziałów ufności dla obu odchyleń standardowych. Dla $\alpha = 1 - p = 1 - 0,95 = 0,05$ z tablic kwantylów rozkładu normalnego odczytujemy stosowny kwantyl:

$$z\left(1 - \frac{\alpha}{2}\right) = z(0,975) = 1,96.$$

Wynik estymacji przedziałowej dla odchyleń standardowych mierzonego ciśnienia w przypadku obu barometrów jest następujący:

$$\sigma_1 \in \left[\frac{\hat{\sigma}_1}{1 + \frac{z\left(1 - \frac{\alpha}{2}\right)}{\sqrt{2 \cdot k}}}, \frac{\hat{\sigma}_1}{1 - \frac{z\left(1 - \frac{\alpha}{2}\right)}{\sqrt{2 \cdot k}}} \right] = \left[\frac{0,8}{1 + \frac{1,96}{\sqrt{2 \cdot 79}}}, \frac{0,8}{1 - \frac{1,96}{\sqrt{2 \cdot 79}}} \right] = [0,69; 0,95] \text{ [Pa]},$$

$$\sigma_2 \in \left[\frac{\hat{\sigma}_2}{1 + \frac{z\left(1 - \frac{\alpha}{2}\right)}{\sqrt{2 \cdot k}}}, \frac{\hat{\sigma}_2}{1 - \frac{z\left(1 - \frac{\alpha}{2}\right)}{\sqrt{2 \cdot k}}} \right] = \left[\frac{0,7}{1 + \frac{1,96}{\sqrt{2 \cdot 79}}}, \frac{0,7}{1 - \frac{1,96}{\sqrt{2 \cdot 79}}} \right] = [0,60; 0,83] \text{ [Pa]}.$$

Widać zatem, że na poziomie ufności 0,95 dokładności barometrów możemy uznać za bardzo zbliżone, gdyż wyznaczone przedziały nie wykluczają się oraz estymatory z obu prób losowych $\hat{\sigma}_1$, $\hat{\sigma}_2$ należą do każdego z przedziałów.

PRZYKŁAD II.4.6

Wynik estymacji punktowej dla wariancji pewnej wielkości, na podstawie pomiarów jednakowo dokładnych, jest następujący: $\sigma^2 \approx 6,5 \pm 2,0 [j^2]$. Odpowiada on wynikowi estymacji przedziałowej na poziomie ufności 0,68 (por. przykład I.2.12 w podrozdziale I.2.5). Wykonać dla tej wariancji i dla odchylenia standardowego estymację przedziałową na poziomie ufności 0,90.

Rozwiązanie

W zadaniu nie podano liczności próby losowej, na podstawie której wykonano estymację punktową. Zatem aby się dowiedzieć, który z modeli należy zastosować do estymacji przedziałowej, należy najpierw oszacować n . Wiemy, że wykonane pomiary były jednakowo dokładne. Możemy przyjąć, że wynik estymacji punktowej powstał na podstawie zastosowania estymatora o lepszych własnościach dla obserwacji jednakowo dokładnych:

$$\sigma^2 = \hat{S}^2 \pm \sigma(\hat{S}^2),$$

$$\hat{\sigma}^2 = \hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x})^2, \quad \sigma(\hat{S}^2) = \sqrt{\frac{2}{n-1}} \cdot \hat{S}^2, \quad \text{z tabeli II.4.3.}$$

Z drugiego wzoru możemy wyliczyć n :

$$\sigma(\hat{S}^2) = \sqrt{\frac{2}{n-1}} \cdot \hat{S}^2 \Leftrightarrow 2,0 = \sqrt{\frac{2}{n-1}} \cdot 6,5 \Rightarrow n \approx 22.$$

Widzimy teraz, że do wykonania estymacji przedziałowej wariancji należy wykorzystać model I. Z tablic kwantyliów rozkładu χ^2 odczytujemy odpowiednie kwantyle dla $\alpha = 1 - p = 1 - 0,90 = 0,10$ przy liczbie stopni swobody $k = n - 1 = 22 - 1 = 21$:

$$\chi^2\left(1 - \frac{\alpha}{2}, k\right) = \chi^2(0,95; 21) = 32,671, \quad \chi^2\left(\frac{\alpha}{2}, k\right) = \chi^2(0,05; 21) = 11,591.$$

Wynikiem estymacji przedziałowej dla wariancji mierzonej wielkości i dla odchylenia standardowego są następujące przedziały:

$$\sigma^2 \in \left[\frac{k \cdot \hat{S}^2}{\chi^2\left(1 - \frac{\alpha}{2}, k\right)}, \frac{k \cdot \hat{S}^2}{\chi^2\left(\frac{\alpha}{2}, k\right)} \right] = \left[\frac{21 \cdot 6,5}{32,671}, \frac{21 \cdot 6,5}{11,591} \right] = [4,18; 11,78] [j^2],$$

$$\sigma \in [\sqrt{4,18}; \sqrt{11,78}] = [2,0; 3,4] [j].$$

II.4.4.3. Estymacja przedziałowa dla wskaźnika struktury

Estymator wskaźnika struktury, zwanego frakcją, został określony w tabeli II.4.3. Podobnie jak dla wariancji – w estymacji przedziałowej tego parametru wyróżniamy dwa modele, w zależności od liczności próby losowej.

Model I

ZAŁOŻENIA

X – ma rozkład dwupunktowy (zero-jedynkowy) z parametrem p ,

n – małe ($n < 100$),

k – liczba wyróżnionych elementów próby losowej (mających wyróżnioną cechę).

Przedział ufności postaci (θ_1, θ_2) spełnia następującą zależność:

$$P(\theta_1(k, n, \alpha) < p < \theta_2(k, n, \alpha)) = 1 - \alpha \quad (\text{II.4.23})$$

Stosowanie efektywnych wzorów na końce θ_1, θ_2 powyższego przedziału ufności, jest dosyć skomplikowane. Wartości te ujęto w tablice dla małej liczności próby, przy poziomie ufności $p = 0,95$. Odpowiednie wartości, zależne od k oraz $(n - k)$ zawiera tablica 9 (dodatek).

Model II (na podstawie twierdzenia II.3.8)

ZAŁOŻENIA

X – ma rozkład dwupunktowy (zero-jedynkowy) z parametrem p ,

n – duże ($n \geq 100$),

k – liczba wyróżnionych elementów próby losowej (posiadających wyróżnioną cechę).

Postać przedziału ufności:

$$[A \cdot (B - C) < p < A \cdot (B + C)] \quad (\text{II.4.24})$$

gdzie:

$$A = \frac{n}{n + z^2 \left(1 - \frac{\alpha}{2}\right)}, \quad B = \frac{k}{n} + \frac{z \left(1 - \frac{\alpha}{2}\right)}{2 \cdot n}, \quad C = \frac{1}{n} \cdot \sqrt{\frac{k(n-k)}{n} + \frac{z^2 \left(1 - \frac{\alpha}{2}\right)}{4}} \quad (\text{II.4.25})$$

$z \left(1 - \frac{\alpha}{2}\right)$ – kwantyl standaryzowanego rozkładu normalnego.

PRZYKŁAD II.4.7

Z populacji 10 000 studentów uczelni technicznej, pięciuset losowo wybranym zadano pytanie, czy w ciągu ostatniego miesiąca przeczytali jakąś książkę beletrystyczną lub o tematyce humanistycznej. Twierdząco odpowiedziało 97 osób. Wyznaczyć przedział ufności dla odsetka osób czytających literaturę inną niż fachowa techniczna na tej uczelni, na poziomie ufności 0,90.

Rozwiązanie

Ze względu na dużą licznosc próby losowej do estymacji przedziałowej wskaźnika struktury zastosujemy model II. Stosowny kwantyl rozkładu normalnego odczytujemy z tablic:

$$p = 0,90; \quad \alpha = 1 - p = 0,10; \quad z\left(1 - \frac{\alpha}{2}\right) = z(0,95) = 1,644854.$$

Następnie wyliczamy parametry definiujące przedział ufności:

$$A = \frac{n}{n + z^2\left(1 - \frac{\alpha}{2}\right)} = \frac{500}{500 + (1,644854)^2} = 0,994618,$$

$$B = \frac{k}{n} + \frac{z\left(1 - \frac{\alpha}{2}\right)}{2 \cdot n} = \frac{97}{500} + \frac{1,644854}{2 \cdot 500} = 0,195645,$$

$$C = \frac{1}{n} \cdot \sqrt{\frac{k(n-k)}{n} + \frac{z^2\left(1 - \frac{\alpha}{2}\right)}{4}} = \frac{1}{500} \cdot \sqrt{\frac{97 \cdot (500 - 97)}{500} + \frac{(1,644854)^2}{4}} = 0,017760.$$

Przedział ufności na poziomie ufności 90% dla frakcji studentów czytających literaturę nietechniczną jest następujący:

$$p \in [A \cdot (B - C); A \cdot (B + C)] = [0,18; 0,21].$$

Oznacza to, że tylko około 18–21% studentów tej uczelni poświęca czas lekturze innej niż związana z ich kierunkiem studiów. ■

II.5. Testowanie hipotez statystycznych

Jak zaznaczono we wstępie rozdziału dotyczącego zagadnień estymacji, testowanie hipotez statystycznych jest drugą, najważniejszą metodą wnioskowania statystycznego. **Hipoteza statystyczna** to każde twierdzenie, które dotyczy nieznanego rozkładu zmiennej losowej. Prawdziwość tego twierdzenia weryfikuje się, opierając się na próbie losowej, za pomocą **testów statystycznych**.

Rozróżniamy dwa rodzaje testów statystycznych:

- 1) **testy parametryczne**,
- 2) **testy nieparametryczne**.

W testach parametrycznych sprawdzane hipotezy dotyczą wartości parametrów rozkładu, a w nieparametrycznych – typu rozkładu, jego kształtu.

W praktyce weryfikacji hipotez, oprócz weryfikowanej hipotezy H_0 , zwanej **hipotezą zerową**, formułuje się również tzw. **hipotezę alternatywną** H_1 . Hipoteza alternatywna jest twierdzeniem, które najbardziej jesteśmy skłonni uznać za prawdziwe, gdy hipotezę zerową trzeba odrzucić. Samo zaś odrzucanie lub przyjęcie hipotezy zerowej odbywa się na podstawie wartości odpowiednio dobranej statystyki, zwanej **funkcją testową**. Jeżeli wartość statystyki znajdzie się w tak zwanym **obszarze przyjęć** hipotezy zerowej, wówczas przyjmujemy hipotezę zerową. Jeżeli natomiast wartość statystyki wpadnie do **obszaru odrzuceń** W , zwanego **obszarem krytycznym** – odrzucamy hipotezę zerową, a przyjmujemy hipotezę alternatywną. Granica pomiędzy obszarami przyjęć i odrzuceń hipotezy wynika z teoretycznej postaci rozkładu statystyki i nazywa się **wartością krytyczną testu**.

Wnioskowanie statystyczne, oparte na wynikach w próbie losowej, jest zawsze obarczone ryzykiem podjęcia błędnej decyzji. Błędna decyzja może przybierać dwojaką formę:

- 1) **błędu pierwszego rodzaju**, oznaczanego α , polegającego na odrzuceniu hipotezy zerowej, gdy w rzeczywistości jest ona prawdziwa,
- 2) **błędu drugiego rodzaju**, oznaczanego β , polegającego na przyjęciu hipotezy zerowej, gdy w rzeczywistości jest ona fałszywa.

Najlepszym rozwiązaniem byłoby jednoczesne zminimalizowanie obu rodzajów błędów. Jednakże nie jest to możliwe. W praktyce postępuje się następująco: arbitralnie ustalamy małe prawdopodobieństwo popełnienia błędu pierwszego rodzaju α , nazywanego **poziomem istotności** (z reguły przyjmuje się jedną z wartości $\alpha = 0,01$, $\alpha = 0,05$, $\alpha = 0,10$). Następnie wyznaczamy taki obszar krytyczny, by w przypadku prawdziwości hipotezy zerowej, prawdopodobieństwo znalezienia się wartości statystyki w tym obszarze wynosiło α . Test, który przy ustalonym prawdopodobieństwie błędu pierwszego rodzaju minimalizuje błąd drugiego rodzaju, nazywamy testem najmocniejszym.

Dla uproszczenia całego zagadnienia, bardzo często nie oblicza się błędu drugiego rodzaju, a wnioski testowe formułuje się następująco:

- „odrzucaamy hipotezę zerową na korzyść hipotezy alternatywnej” – jeżeli wartość statystyki wpada do obszaru krytycznego,
- „nie ma podstaw do odrzucenia hipotezy zerowej” – jeżeli wartość statystyki nie wpada do obszaru krytycznego.

Taka konstrukcja testu jest znacznie prostsza, ale jej zasadniczą wadą jest to, że nigdy nie podejmujemy decyzji o zdecydowanym przyjęciu hipotezy zerowej, ponieważ odrzucamy możliwość popełnienia błędu drugiego rodzaju. Ten typ testów, o uproszczonej konstrukcji, nazywa się **testami istotności**. I właśnie ten rodzaj testów zostanie omówiony w niniejszym rozdziale.

Podsumowując, możemy napisać, że konstrukcja testu przebiega według następujących etapów:

1. Postawienie hipotezy do weryfikacji H_0 .
2. Wybór statystyki (funkcji testowej), stosownie do treści hipotezy zerowej H_0 oraz do warunków spełnianych przez próbę losową.
3. Ustalenie poziomu istotności testu α .
4. Określenie hipotezy alternatywnej H_1 , na podstawie wyników z próby losowej.
5. Ustalenie granic obszaru krytycznego W , stosownie do treści hipotezy alternatywnej H_1 tak, by jego pole powierzchni było równe poziomowi istotności α .
6. Wnioskowanie na podstawie położenia wartości statystyki względem obszaru krytycznego.

Ustalając poziom istotności α , możemy kierować mocą testu. Wniosek o odrzuceniu hipotezy zerowej jest tym bardziej wiarygodny, im α jest mniejsze. Podobnie bardziej ufamy wnioskowi o braku podstaw do odrzucenia H_0 dla większego α .

Kolejność zaprezentowanych dalej testów w podrozdziałach II.5.1.3 i II.5.1.4, na pozór nie odpowiadająca kolejności podrozdziałów II.5.1.1 i II.5.1.2, wynika z faktu, iż w testach na porównanie dwóch wartości przeciętnych (podrozdz. II.5.1.4) może zachodzić konieczność wcześniejszego przetestowania równości wariancji w tych samych próbach losowych (podrozdz. II.5.1.3).

II.5.1. Testy parametryczne

Jak wspomniano we wstępie, testy parametryczne dotyczą weryfikacji wartości parametrów nieznanego rozkładu zmiennej losowej. Najczęściej testowanymi parametrami są wartość przeciętna oraz wariancja. W niniejszym rozdziale omówiono testy dotyczące tych dwóch parametrów oraz wskaźnika struktury.

II.5.1.1. Testowanie pojedynczej wartości przeciętnej

Prosta hipoteza zerowa do weryfikacji wartości przeciętnej analizowanej zmiennej losowej ma następującą postać:

$$H_0: \mu = \mu_0 \quad (\text{II.5.1})$$

czyli stawiamy hipotezę, że wartość przeciętna μ jest równa ustalonej wartości μ_0 . Wartość μ_0 może wynikać np. z wcześniejszych pomiarów lub danych podawanych przez producenta określonego wyrobu.

Możliwe są następujące postacie hipotezy alternatywnej:

$$H_1: \mu > \mu_0 \quad (\text{II.5.2a})$$

$$H_1: \mu < \mu_0 \quad (\text{II.5.2b})$$

$$H_1: \mu \neq \mu_0 \quad (\text{II.5.2c})$$

W praktyce hipoteza alternatywna to twierdzenie, które chcemy udowodnić, wynikające z wyników doświadczenia i celu przeprowadzanych badań statystycznych.

W zależności od posiadanych informacji na temat badanej populacji i w zależności od liczności próby losowej pobranej do analiz x_1, x_2, \dots, x_n , wyróżniamy trzy modele testu pojedynczej wartości przeciętnej. Modele te, w swoich założeniach, odpowiadają modelom z zagadnienia estymacji przedziałowej dla tego parametru.

Model I (na podstawie twierdzenia I.2.1)

ZAŁOŻENIA

$X \sim N(\mu, \sigma)$,

σ – znane,

n – dowolne.

Postać funkcji testowej (statystyki):

$$Z = \frac{\hat{X} - \mu_0}{\sigma(\hat{X})} = \frac{\hat{X} - \mu_0}{\sigma / \sqrt{n}} = \sqrt{n} \cdot \frac{\hat{X} - \mu_0}{\sigma} \quad (\text{II.5.3})$$

Zmienna Z ma standaryzowany rozkład normalny $N(0,1)$, a zatem ten rozkład wykorzystamy do konstrukcji obszaru krytycznego.

Dla każdej z trzech hipotez alternatywnych obszar krytyczny W przyjmuje inną postać. Dla pierwszej z nich (II.5.2a) będzie to tzw. prawostronny obszar krytyczny, ograniczony kwantylem rozkładu normalnego $[z(1 - \alpha), +\infty)$ – rysunek II.5.1a.

Dla drugiej hipotezy alternatywnej (II.5.2b) mamy lewostronny obszar krytyczny, ograniczony kwantylem $z(\alpha)$ rozkładu normalnego $(-\infty, z(\alpha)]$ – rysunek II.5.1b.

A dla trzeciej (II.5.2c) mamy dwustronny obszar krytyczny, będący sumą dwóch przedziałów $\left(-\infty, z\left(\frac{\alpha}{2}\right)\right] \cup \left[z\left(1 - \frac{\alpha}{2}\right), +\infty\right)$ – rysunek II.5.1c.

Oczywiste jest, że:

$$z(1 - \alpha) = -z(\alpha) \quad \text{oraz} \quad z\left(1 - \frac{\alpha}{2}\right) = -z\left(\frac{\alpha}{2}\right) \quad (\text{II.5.4})$$

Model II (na podstawie twierdzenia II.3.7)

ZAŁOŻENIA

$X \sim N(\mu, \sigma)$,

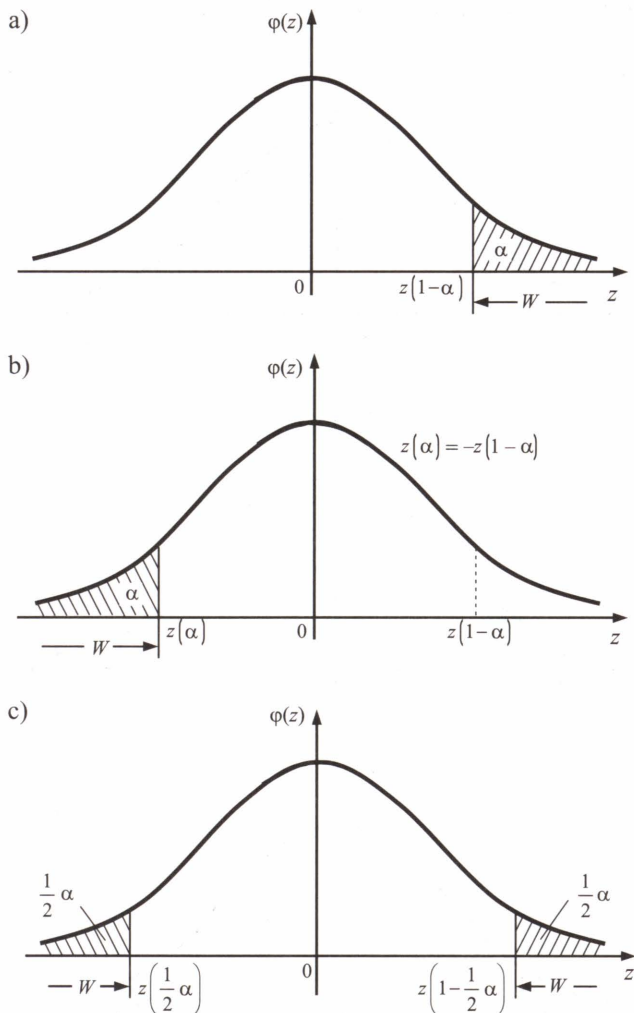
σ – nieznane,

n – dowolne (w szczególności małe $n < 30$).

Postać funkcji testowej (statystyki):

$$T = \frac{\hat{X} - \mu_0}{\sigma(\hat{X})} \quad (\text{II.5.5})$$

gdzie: $\sigma(\hat{X}) = \frac{\hat{S}}{\sqrt{n}}$ lub $\sigma(\hat{X}) = \frac{\bar{S}}{\sqrt{n-1}}$, \hat{S} i \bar{S} są określone wzorami (II.3.3) i (II.3.1).

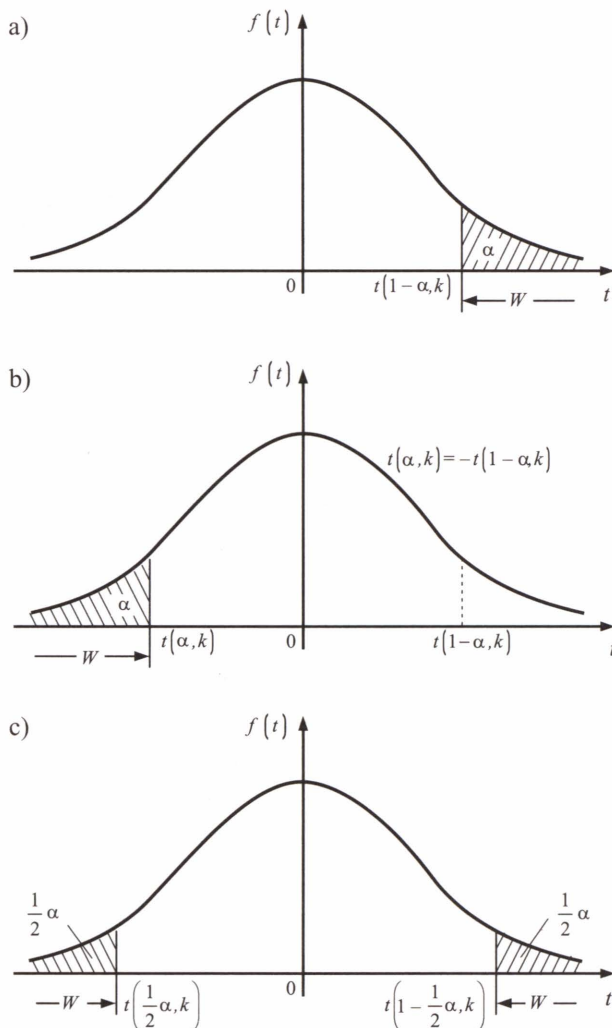


Rys. II.5.1. Obszary krytyczne w teście wartości przeciętnej – model I: a) prawostronny; b) lewostronny; c) dwustronny. α – poziom istotności, $z(1 - \alpha)$, $z(\alpha)$, $z\left(\frac{\alpha}{2}\right)$, $z\left(1 - \frac{\alpha}{2}\right)$ – kwantyle standaryzowanego rozkładu normalnego

Zmienna T ma rozkład T-Studenta o $k = n - 1$ stopniach swobody. Ten rozkład wykorzystamy do konstrukcji obszaru krytycznego, który powstaje analogicznie jak w modelu I, dla poszczególnych hipotez alternatywnych (rys. II.5.2a–c), ale przy uwzględnieniu stosownych kwantylów rozkładu T-Studenta:

$$t(1-\alpha, k), t(\alpha, k), t\left(\frac{\alpha}{2}, k\right), t\left(1-\frac{\alpha}{2}, k\right)$$

między którymi zachodzą analogiczne relacje do (II.5.4).



Rys. II.5.2. Obszary krytyczne w teście wartości przeciętnej – model II: a) prawostronny; b) lewostronny; c) dwustronny

Model III (na podstawie twierdzenia II.3.2)

ZAŁOŻENIA

X – ma dowolny rozkład,

σ – nieznane, skończone,

n – duże ($n > 30$).

Postać statystyki oraz obszar krytyczny są analogiczne jak w modelu I, przy czym odchylenie standardowe σ estymujemy za pomocą odchylenia \hat{S} lub \bar{S} wyznaczonego z próby losowej x_1, x_2, \dots, x_n .

Oczywiście pole powierzchni obszaru krytycznego równa się poziomowi istotności α . Gdy wartość statystyki należy do obszaru krytycznego, hipotezę zerową $H_0: \mu = \mu_0$ należy odrzucić. W sytuacji przeciwnej nie ma podstaw do jej odrzucenia na zadanym poziomie istotności.

PRZYKŁAD II.5.1

Sieć sklepów spożywczych zaopatrujących się w tych samych zakładach mleczarskich wątpi w uczciwość dostawcy jeśli chodzi o zawartość tłuszczu w produktach mlecznych. Podejrzewa się, iż zawartość ta jest mniejsza od podawanej na opakowaniu. Jednocześnie nie ma podstaw, by sądzić, że odchylenie standardowe zawartości tłuszczu jest inne niż deklarowane przez producenta, równe 0,3%. W celu zweryfikowania podejrzeń wykonano stosowny test statystyczny, zakładając normalność rozkładu procentowej zawartości tłuszczu w badanych produktach. Z 20 różnych partii towaru wylosowano po jednym opakowaniu śmietany 18-procentowej, uzyskując następujące wartości analizowanej zmiennej losowej (procentowej zawartości tłuszczu):

18,2; 17,0; 17,8; 17,9; 16,9; 18,4; 16,8; 16,9; 17,4; 19,0;

16,9; 17,3; 17,7; 18,1; 16,8; 16,7; 17,7; 17,5; 17,9; 16,9.

Rozwiązanie

Weryfikowana hipoteza jest następującej postaci:

$$H_0: \mu = 18,$$

przy czym μ oznacza średnią procentową zawartość tłuszczu w śmietanie.

Hipotezę alternatywną, zgodnie z podejrzeniami handlowców, formułujemy następująco:

$$H_0: \mu < 18,$$

Ze względu na znane odchylenie standardowe, do przeprowadzenia testu wykorzystano model I.

Średnia w próbie losowej wynosi 17,49, a zatem funkcja testowa (II.5.3) przyjmuje wartość:

$$Z = \sqrt{n} \cdot \frac{\hat{X} - \mu_0}{\sigma} = \sqrt{20} \frac{17,49 - 18}{0,3} = -7,60.$$

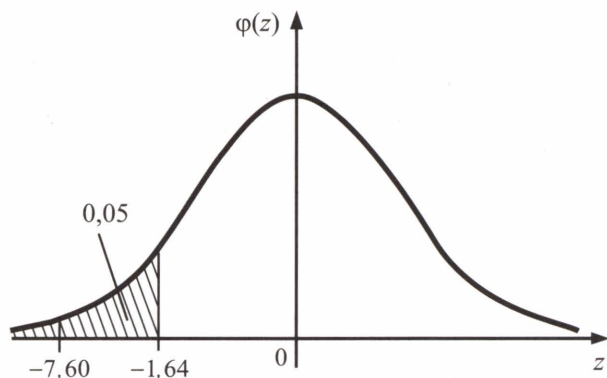
Kwantyle, pełniące rolę wartości krytycznych, postaci $z(\alpha)$ dla różnych poziomów istotności α , wynoszą:

$$z(0,01) = -z(0,99) = -2,33,$$

$$z(0,05) = -z(0,95) = -1,64,$$

$$z(0,10) = -z(0,90) = -1,28.$$

Wartość statystyki jest zdecydowanie mniejsza od każdej z powyższych wartości krytycznych, zatem mieści się w obszarze krytycznym $(-\infty, z(\alpha))$ (rys. II.5.3), czyli hipotezę zerową H_0 należy odrzucić na korzyść hipotezy alternatywnej, na każdym z rozpatrywanych poziomów istotności α . Innymi słowy, podejrzenia o nieuczciwość producenta okazały się słuszne: normy zawartości tłuszczu w śmietanie są zaniżane.



Rys. II.5.3. Lewostronny obszar krytyczny w teście średniej zawartości tłuszczu dla poziomu istotności $\alpha = 0,05$

Warto zwrócić uwagę na fakt, że odrzucenie hipotezy zerowej nie powinno prowadzić do podejmowania pochopnych wniosków w sposób automatyczny. W szczególności nie należy mylić istotności statystycznej z praktyczną istotnością. Rozbieżności w tym zakresie mają miejsce, gdy zastosowany test jest zbyt czuły, tzn. gdy jego moc jest zbyt duża: zbyt duże α , powodujące odrzucenie H_0 , gdy wartość statystyki jest bliska wartości krytycznej.

W przykładzie II.5.1 średnia policzona z próby (17,49) jest wyraźnie mniejsza od wartości nominalnej (18) w kontekście odchylenia standardowego (0,3). Zatem można wnioskować, że otrzymany wynik jest istotny zarówno statystycznie, jak i praktycznie, czyli producent rzeczywiście zaniża normy produkcyjne.

II.5.1.2. Testowanie pojedynczej wariancji

Prosta hipoteza zerowa do weryfikacji wariancji analizowanej zmiennej losowej ma następującą postać:

$$H_0: \sigma^2 = \sigma_0^2 \quad (\text{II.5.6})$$

Możliwe postacie hipotezy alternatywnej:

$$H_1: \sigma^2 > \sigma_0^2 \quad (\text{II.5.7a})$$

$$H_1: \sigma^2 < \sigma_0^2 \quad (\text{II.5.7b})$$

$$H_1: \sigma^2 \neq \sigma_0^2 \quad (\text{II.5.7c})$$

W zależności od liczności próby statystycznej x_1, x_2, \dots, x_n , wyróżniamy dwa modele testu pojedynczej wariancji. Modele te, w swoich założeniach, odpowiadają modelom z zagadnienia estymacji przedziałowej dla tego parametru.

Model I (na podstawie twierdzeń II.3.3 lub II.3.5)

ZAŁOŻENIA

$$X \sim N(\mu, \sigma),$$

μ, σ – nieznane,

n – dowolne (w szczególności małe $n < 50$).

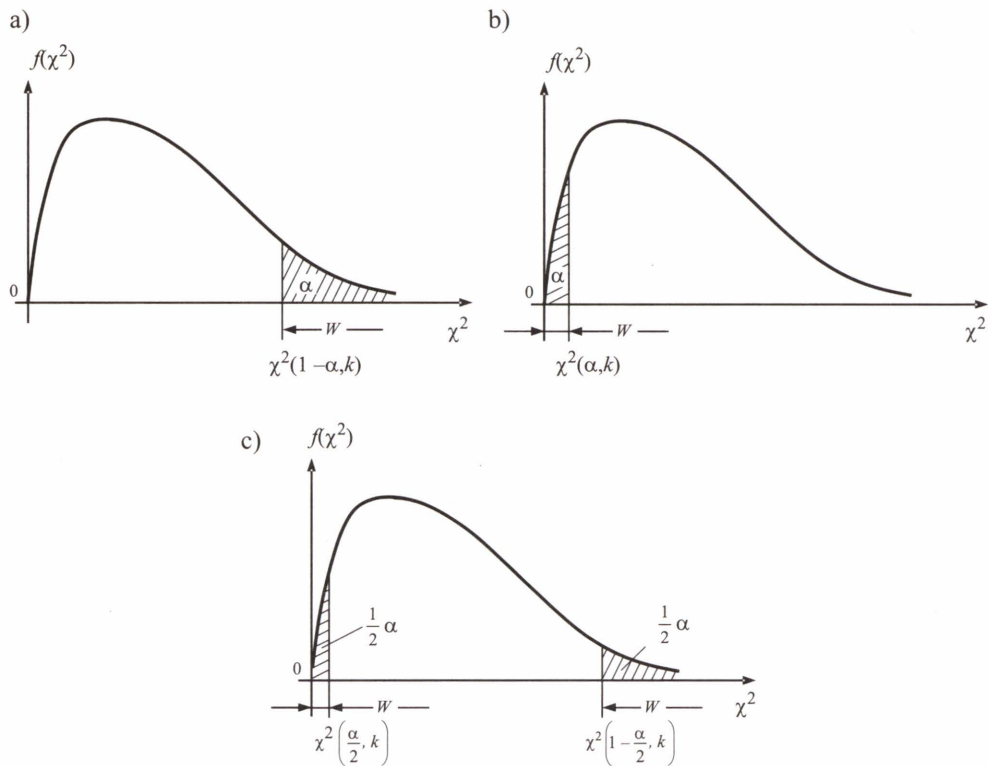
Postać statystyki:

$$\chi^2 = \frac{(n-1) \cdot \hat{S}^2}{\sigma_0^2} \quad \text{lub} \quad \chi^2 = \frac{n \cdot \bar{S}^2}{\sigma_0^2} \quad (\text{II.5.8})$$

gdzie: \hat{S}^2 i \bar{S}^2 są określone wzorami (II.3.3) i (II.3.1).

Zmienna χ^2 ma rozkład chi-kwadrat o $k = n - 1$ stopniach swobody. Rozkład ten wykorzystamy do konstrukcji obszaru krytycznego. Dla każdej z trzech hipotez alternatywnych obszar krytyczny przyjmuje inną postać. Dla pierwszej z nich (II.5.7a) będzie to prawostronny obszar krytyczny, ograniczony kwantylem rozkładu chi-kwadrat $[\chi^2(1 - \alpha, k), +\infty)$ – rysunek II.5.4a. Dla drugiej hipotezy alternatywnej (II.5.7b) mamy lewostronny obszar krytyczny, ograniczony kwantylem $\chi^2(\alpha, k)$ rozkładu chi-kwadrat $[-\infty, \chi^2(\alpha, k))$ – rysunek II.5.4b.

A dla trzeciej mamy dwustronny obszar krytyczny, będący sumą dwóch przedziałów $\left(0, \chi^2\left(\frac{\alpha}{2}, k\right)\right] \cup \left[\chi^2\left(1 - \frac{\alpha}{2}, k\right), +\infty\right)$ – rysunek II.5.4c.



Rys. II.5.4. Obszary krytyczne w teście wariancji: a) prawostronny; b) lewostronny; c) dwustronny

Ze względu na niesymetryczność rozkładu chi-kwadrat, dwustronny obszar krytyczny (rys. II.5.4c) również nie jest położony symetrycznie względem wartości przeciętnej zmiennej losowej χ^2 , dla której funkcja gęstości $f(\chi^2)$ osiąga maksimum.

Model II (na podstawie twierdzenia II.3.2. Lindeberga–Levy’ego)

ZAŁOŻENIA

$X \sim N(\mu, \sigma)$,

σ – nieznane,

n – duże ($n > 50$).

Postać statystyki:

$$Z = \sqrt{2 \cdot \chi^2 - \sqrt{2n - 3}} \quad (\text{II.5.9})$$

gdzie χ^2 to statystyka z modelu I, określona wzorem (II.5.8).

Zmienna Z ma rozkład normalny $N(0,1)$ (patrz [6]), więc konstrukcja obszaru krytycznego przebiega jak w modelu I lub III dla pojedynczej wartości przeciętnej.

W każdym przypadku pole powierzchni obszaru krytycznego równa się poziomowi istotności α . Gdy wartość statystyki mieści się w obszarze krytycznym, hipotezę zerową $H_0: \sigma^2 = \sigma_0^2$ należy odrzucić. W sytuacji przeciwnej nie ma podstaw do jej odrzucenia na danym poziomie istotności.

PRZYKŁAD II.5.2

W celu zweryfikowania nominalnej dokładności przyrządu pomiarowego równej 2 mm, wykonano serię 55 pomiarów, uzyskując oszacowanie wariancji pomiaru 4,7 mm². Za pomocą stosownego testu statystycznego sprawdzić na poziomie istotności $\alpha = 0,15$, czy badany przyrząd faktycznie mierzy z podaną przez producenta dokładnością.

Rozwiązanie

Na podstawie wartości wariancji z próby wyników pomiaru wiemy, że oszacowane z próby, za pomocą estymatora nieobciążonego, odchylenie standardowe wynosi:

$$\hat{\sigma} = \hat{S} = \sqrt{4,7 \text{ mm}^2} = 2,17 \text{ mm},$$

a zatem jest ono nieznacznie większe od deklarowanego przez producenta przyrządu, czyli dokładność pomiaru jest nieznacznie mniejsza. Chcemy sprawdzić, czy ta różnica jest statystycznie istotna.

Hipoteza zerowa odpowiedniego testu parametrycznego jest postaci (II.5.6):

$$H_0: \sigma^2 = 2^2.$$

Ze względu na niewielką różnicę pomiędzy teoretyczną (4,0 mm²) i doświadczalną (4,7 mm²) wartością wariancji (różnica ta może być dziełem przypadku), hipoteza alternatywna będzie postaci (II.5.7c):

$$H_1: \sigma^2 \neq 2^2.$$

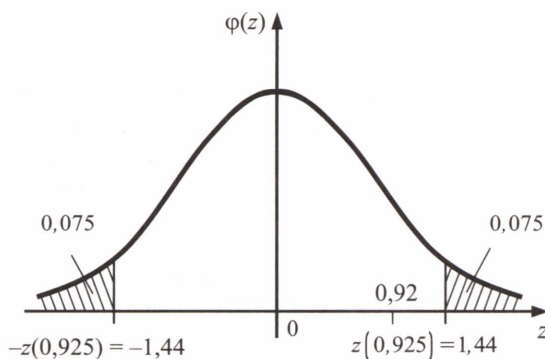
Do weryfikacji wykorzystamy model II, gdyż próba losowa jest odpowiednio duża $n = 55$. Zatem wartość statystyki (II.5.9) wynosi:

$$Z = \sqrt{2 \cdot \chi^2} - \sqrt{2n-3} = \sqrt{2 \cdot \frac{(n-1) \cdot \hat{S}^2}{\sigma_0^2}} - \sqrt{2n-3} = \sqrt{2 \cdot \frac{(55-1) \cdot 4,7}{4}} - \sqrt{2 \cdot 55 - 3} = 0,92.$$

Dla ustalonego poziomu istotności $\alpha = 0,15$, wobec postaci hipotezy alternatywnej H_1 , mamy dwustronny obszar krytyczny będący sumą następujących przedziałów:

$$\begin{aligned} \left(-\infty; z\left(\frac{\alpha}{2}\right)\right] \cup \left[z\left(1-\frac{\alpha}{2}\right); +\infty\right) &= \left(-\infty; -z\left(1-\frac{\alpha}{2}\right)\right] \cup \left[z\left(1-\frac{\alpha}{2}\right); +\infty\right) = \\ &= (-\infty; -z(0,925)] \cup [z(0,925); +\infty) = (-\infty; -1,44] \cup [1,44; +\infty). \end{aligned}$$

Jak widać z rysunku II.5.5, wartość funkcji testowej nie należy do obszaru krytycznego, a zatem nie ma podstaw do odrzucenia hipotezy zerowej, zakładającej, że rzeczywista dokładność przyrządu jest równa nominalnej.



Rys. II.5.5. Dwustronny obszar krytyczny w teście weryfikującym dokładność przyrządu – model II

Sprawdźmy, czy powyższy wniosek potwierdzi się w sytuacji sformułowania hipotezy alternatywnej w postaci (II.5.7a):

$$H_1: \sigma^2 > 2^2.$$

Prawostronny obszar krytyczny będzie w tym przypadku następującej postaci:

$$[z(1 - \alpha); +\infty) = [z(0,85); +\infty) = [1,04; +\infty).$$

Również i tym razem wartość statystyki Z należy do obszaru przyjęć hipotezy zerowej. Zatem wniosek będzie identyczny z poprzednim.

Taki sam wynik dla mocniejszego testu zwiększa zaufanie do postawionego wniosku. ■

II.5.1.3. Test porównujący dwie wariancje

Gdy prowadzimy badania ze względu na pewną cechę w dwóch populacjach, często zachodzi konieczność porównania rozproszenia wartości tej cechy w obu populacjach. W tym celu wykonujemy test porównujący dwie wariancje, dla którego hipoteza zerowa ma postać:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad (\text{II.5.10})$$

Możliwe postacie hipotezy alternatywnej:

$$H_1: \sigma_1^2 > \sigma_2^2 \quad (\text{II.5.11a})$$

$$H_1: \sigma_1^2 < \sigma_2^2 \quad (\text{II.5.11b})$$

$$H_0: \sigma_1^2 \neq \sigma_2^2 \quad (\text{II.5.11c})$$

Model (na podstawie twierdzenia II.3.6)

ZAŁOŻENIA

Badana cecha ma w dwóch populacjach rozkłady normalne:

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ o nieznanych parametrach,

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ – próby losowe.

Postać statystyki:

$$F = \frac{\hat{S}_1^2}{\hat{S}_2^2} \quad (\text{II.5.12})$$

gdzie: \hat{S}_1^2, \hat{S}_2^2 – wariancje z prób losowych odpowiednio x_1, x_2, \dots, x_n oraz y_1, y_2, \dots, y_m .

Zmienna F ma rozkład F-Snedecora o $(n-1, m-1)$ stopniach swobody. W praktyce, ze względu na konstrukcję tablic wartości krytycznych tego rozkładu (wszystkie wartości są większe od 1) w statystyce F do licznika wstawiamy estymator wariancji o większej wartości. Konstrukcja obszaru krytycznego przebiega podobnie jak w innych testach.

II.5.1.4. Testy porównujące dwie wartości przeciętne

Porównanie dwóch populacji ze względu na wybraną cechę to przede wszystkim porównanie średnich wartości tej cechy w obu populacjach. W tym celu stosujemy testy na porównanie dwóch wartości przeciętnych. Hipoteza zerowa w tych testach jest analogiczna do hipotezy porównującej dwie wariancje:

$$H_0: \mu_1 = \mu_2 \quad (\text{II.5.13})$$

Możliwe postacie hipotezy alternatywnej:

$$H_0: \mu_1 > \mu_2 \quad (\text{II.5.14a})$$

$$H_0: \mu_1 < \mu_2 \quad (\text{II.5.14b})$$

$$H_0: \mu_1 \neq \mu_2 \quad (\text{II.5.14c})$$

Model I

ZAŁOŻENIA

Badana cecha ma w dwóch populacjach rozkłady normalne:

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ o znanych odchyleniach standardowych σ_1, σ_2 i nieznanych wartościach przeciętnych μ_1, μ_2 ,

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ – próby losowe.

Postać funkcji testowej:

$$Z = \frac{\hat{X} - \hat{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \quad (\text{II.5.15})$$

Statystyka powyższa, przy założeniu prawdziwości hipotezy zerowej, ma standaryzowany rozkład normalny $N(0,1)$. Zatem dla poszczególnych hipotez alternatywnych (II.5.14a–c), obszarami krytycznymi testu są zbiory podobne do tych z modelu I dla pojedynczej wartości przeciętnej.

Model II

ZAŁOŻENIA

Badana cecha ma w dwóch populacjach rozkłady normalne:

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ o nieznanach, ale jednakowych wariancjach ($\sigma_1 = \sigma_2$),

μ_1, μ_2 – nieznane,

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ – próby losowe.

Postać funkcji testowej:

$$T = \frac{\hat{X} - \hat{Y}}{\sqrt{\frac{(n-1) \cdot \hat{S}_1^2 + (m-1) \cdot \hat{S}_2^2}{n+m-2} \cdot \frac{n+m}{n \cdot m}}} \quad (\text{II.5.16})$$

Statystyka (II.5.16) przy założeniu prawdziwości hipotezy zerowej, ma rozkład T-Studenta o $(n+m-2)$ stopniach swobody. Zatem dla poszczególnych hipotez alternatywnych (II.5.14a–c), obszarami krytycznymi testu są następujące zbiory:

$$[t(1-\alpha, n+m-2), +\infty) \quad (\text{II.5.17a})$$

$$(-\infty, -t(1-\alpha, n+m-2)] \quad (\text{II.5.17b})$$

$$\left(-\infty, -t\left(1-\frac{\alpha}{2}, n+m-2\right)\right] \cup \left[t\left(1-\frac{\alpha}{2}, n+m-2\right), +\infty\right) \quad (\text{II.5.17c})$$

Założenie o równości wariancji w dwóch populacjach możemy uznać za spełnione, np. w sytuacji, gdy pomiar badanej cechy w obu populacjach wykonuje ten sam obserwator, w jednakowych warunkach zewnętrznych i przy zastosowaniu tego samego przyrządu pomiarowego lub gdy długotrwały pomiar odchyłeń standardowych na podstawie prób losowych, pobieranych z obu populacji w jednakowych odstępach czasu, wskazuje na równość tych odchyłeń. Najczęściej jednak nie jesteśmy w stanie *a priori* stwierdzić, czy wariancje są jednakowe. Dlatego w celu sprawdzenia tego założenia należy najpierw wykonać test na porównanie wariancji w dwóch populacjach, omówiony w podrozdziale II.5.1.3. W przypadku gdy nie będzie podstaw do odrzucenia hipotezy o równości wariancji – możemy zastosować powyższy model testu do zweryfikowania hipotezy o równości wartości przeciętnych. W sytuacji przeciwnej stosujemy test według modelu III.

Model III

ZAŁOŻENIA

Badana cecha ma w dwóch populacjach rozkłady normalne:

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ o nieznanach parametrach,

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ – próby losowe.

Postać funkcji testowej:

$$C = \frac{\hat{X} - \hat{Y}}{\sqrt{\frac{\hat{S}_1^2}{n} + \frac{\hat{S}_2^2}{m}}} \quad (\text{II.5.18})$$

Jest to tzw. statystyka Cochran–Coxa, a jej rozkład zależy od liczności prób losowych oraz od stosunku odchyłeń standardowych z obu populacji σ_1/σ_2 . Stosunku tego nie znamy, ze względu na nieznaną wartość parametrów rozkładu; jednakże dla danych liczności n i m można określić przybliżoną wartość kwantyla rzędu p rozkładu statystyki C :

$$C(p, n, m) \cong \left(\frac{\hat{S}_1^2}{n} t(p, n-1) + \frac{\hat{S}_2^2}{m} t(p, m-1) \right) : \left(\frac{\hat{S}_1^2}{n} + \frac{\hat{S}_2^2}{m} \right) \quad (\text{II.5.19})$$

gdzie: $t(p, n-1)$, $t(p, m-1)$ – kwantyle rozkładu T-Studenta.

Zatem dla poszczególnych hipotez alternatywnych (II.5.14a–c), obszarami krytycznymi testu są następujące zbiory:

$$[c(1 - \alpha, n, m), +\infty) \quad (\text{II.5.20a})$$

$$(-\infty, -c(1 - \alpha, n, m)] \quad (\text{II.5.20b})$$

$$\left(-\infty, -c\left(1 - \frac{\alpha}{2}, n, m\right) \right] \cup \left[c\left(1 - \frac{\alpha}{2}, n, m\right), +\infty \right) \quad (\text{II.5.20c})$$

Model IV

ZAŁOŻENIA

Badana cecha ma w dwóch populacjach rozkłady normalne:

$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$ o nieznanach parametrach,

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ – próby losowe,

n, m – duże (powyżej 100).

Postać statystyki oraz obszar krytyczny są analogiczne jak w modelu I, przy czym odchylenia standardowe σ_1, σ_2 estymujemy za pomocą odchyłeń wyznaczonych z prób losowych: \hat{S}_1, \hat{S}_2 .

PRZYKŁAD II.5.3

Pojemność płuc osób uprawiających czynnie sport ma rozkład normalny o odchyleniu standardowym 440 cm^3 , natomiast u osób nieuprawiających sportu pojemność płuc ma rozkład normalny o odchyleniu standardowym 620 cm^3 . Wylosowano z obu populacji próby o liczebności odpowiednio 20 i 15 osób, z których otrzymano średnie, odpowiednio: 4080 cm^3 i 3610 cm^3 . Na poziomie istotności $\alpha = 1\%$ sprawdzić hipotezę, że uprawianie sportu zwiększa istotnie pojemność życiową płuc.

Rozwiązanie

Należy porównać wartości przeciętne μ_1, μ_2 pojemności płuc osób z dwóch prób losowych, pobranych z dwóch różnych populacji, odpowiednio: uprawiających czynnie sport i nieuprawiających go. Zatem hipoteza zerowa będzie postaci (II.5.13):

$$H_0: \mu_1 = \mu_2.$$

Ze względu na wyniki pomiaru pojemności płuc w obu grupach i jasno określony cel testu, postać hipotezy alternatywnej to:

$$H_1: \mu_1 > \mu_2,$$

przy czym μ_1 odnosi się do osób aktywnych sportowo.

Ponieważ zakłada się normalność rozkładu analizowanej zmiennej losowej oraz znane są odchylenia standardowe w obu populacjach, można zastosować model I, w którym funkcja testowa przyjmuje następującą wartość:

$$Z = \frac{\hat{X} - \hat{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{4080 - 3610}{\sqrt{\frac{440^2}{20} + \frac{620^2}{15}}} = \frac{470}{\sqrt{9680 + 25626,7}} = 2,50.$$

Ze względu na postać H_1 mamy do czynienia z prawostronnym obszarem krytycznym postaci:

$$[z(1 - \alpha); +\infty) = [z(0,99); +\infty) = [2,33; +\infty),$$

do którego należy wartość statystyki Z . Zatem trzeba odrzucić hipotezę zerową na korzyść alternatywnej, czyli test wykazał, że pojemność płuc osób czynnie zajmujących się sportem jest istotnie większa od pojemności płuc preferujących inny sposób spędzania czasu.

■

II.5.1.5. Testowanie wartości wskaźnika struktury

Chcąc zweryfikować hipotezę na temat tego, jaki procent populacji stanowią elementy posiadające wyróżnioną cechę, należy wykonać test na wartość wskaźnika struktury, czyli frakcji.

Prosta hipoteza zerowa do weryfikacji frakcji ma następującą postać:

$$H_0: p = p_0 \quad (\text{II.5.21})$$

Możliwe są następujące postacie hipotezy alternatywnej:

$$H_1: p > p_0 \quad (\text{II.5.22a})$$

$$H_1: p < p_0 \quad (\text{II.5.22b})$$

$$H_1: p \neq p_0 \quad (\text{II.5.22c})$$

Do weryfikacji powyższych hipotez stosujemy jeden z poniższych modeli w zależności od liczebności próby losowej.

Model I

ZAŁOŻENIA

Badana cecha ma rozkład dwupunktowy o nieznanym parametrze p ,
 n – duże ($n \geq 100$).

Postać statystyki:

$$Z = \frac{k - n \cdot p_0}{\sqrt{n \cdot p_0 \cdot (1 - p_0)}} \quad (\text{II.5.23})$$

gdzie k – liczba elementów posiadających wyróżnioną cechę, w próbie o liczebności n .

Przy założeniu prawdziwości hipotezy zerowej, statystyka (II.5.23) ma w przybliżeniu standaryzowany rozkład normalny $N(0,1)$. Przybliżenie jest wystarczająco dobre, gdy $np_0 \geq 50$.

Model II

ZAŁOŻENIA

Badana cecha ma rozkład dwupunktowy o nieznanym parametrze p ,
 n – małe.

Postać statystyki:

$$Z = \left(2 \cdot \arcsin \sqrt{\frac{k}{n}} - 2 \cdot \arcsin \sqrt{p_0} \right) \cdot \sqrt{n} \quad (\text{II.5.24})$$

gdzie k – liczba elementów posiadających wyróżnioną cechę, w próbie o liczebności n .

Przy założeniu prawdziwości hipotezy zerowej oraz dla $k \neq 0$ i $k \neq n$, powyższa statystyka ma w przybliżeniu standaryzowany rozkład normalny $N(0,1)$.

W obu modelach kwantyle standaryzowanego rozkładu normalnego służą do wyznaczenia granic obszaru krytycznego. Zatem dla poszczególnych hipotez alternatywnych (II.5.22a–c), przy poziomie istotności α , obszarami krytycznymi testu są następujące przedziały:

$$[z(1 - \alpha), +\infty) \quad (\text{II.5.25a})$$

$$(-\infty, -z(1-\alpha)] \quad (\text{II.5.25b})$$

$$\left(-\infty, -z\left(1-\frac{\alpha}{2}\right)\right] \cup \left[z\left(1-\frac{\alpha}{2}\right), +\infty\right) \quad (\text{II.5.25c})$$

II.5.1.6. Testy porównujące dwa wskaźniki struktury

Chcąc porównać frakcje elementów mających wyróżnioną cechę w dwóch populacjach, weryfikujemy hipotezę zerową następującej postaci:

$$H_0: p_1 = p_2 \quad (\text{II.5.26})$$

przeciw jednej z następujących hipotez alternatywnych:

$$H_1: p_1 > p_2 \quad (\text{II.5.27a})$$

$$H_1: p_1 < p_2 \quad (\text{II.5.27b})$$

$$H_1: p_1 \neq p_2 \quad (\text{II.5.27c})$$

Model I

ZAŁOŻENIA

Badana cecha ma w obu próbach rozkład dwupunktowy o nieznanach parametrach p_1, p_2 ,

n_1, n_2 – duże (co najmniej po 100 elementów w obu próbach).

Postać statystyki:

$$Z = \frac{\frac{k_1}{n_1} - \frac{k_2}{n_2}}{\sqrt{\frac{\frac{k_1 + k_2}{n_1 + n_2} \cdot \left(1 - \frac{k_1 + k_2}{n_1 + n_2}\right)}{\frac{n_1 \cdot n_2}{n_1 + n_2}}}} \quad (\text{II.5.28})$$

gdzie k_i – liczba elementów wyróżnionych, w próbie o liczności n_i .

Przy założeniu prawdziwości hipotezy zerowej, powyższa statystyka ma asymptotycznie standaryzowany rozkład normalny $N(0,1)$.

Model II

ZAŁOŻENIA

Badana cecha ma w obu próbach rozkład dwupunktowy o nieznanach parametrach p_1, p_2 ,

n_1, n_2 – małe.

Postać statystyki:

$$Z = 2 \cdot \left(\arcsin \sqrt{\frac{k_1}{n_1}} - \arcsin \sqrt{\frac{k_2}{n_2}} \right) \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \quad (\text{II.5.29})$$

gdzie k_i – liczba elementów wyróżnionych, w próbie o liczności n_i .

Przy założeniu prawdziwości hipotezy zerowej, powyższa statystyka ma w przybliżeniu standaryzowany rozkład normalny $N(0,1)$.

W obu modelach obszar krytyczny dla ustalonej hipotezy alternatywnej tworzymy jak w teście pojedynczej wartości wskaźnika struktury. A wnioskowanie statystyczne przebiega jak w każdym innym teście istotności.

PRZYKŁAD II.5.4

W celu porównania stanu dbałości o środowisko naturalne przez zakłady przemysłowe z analogicznymi badaniami przeprowadzonymi w roku poprzednim, wylosowano niezależnie 300 zakładów. W 240 stwierdzono zachowywanie wszystkich norm ochrony środowiska. W roku wcześniejszym na 250 wylosowanych zakładów 175 spełniało ww. normy. Na poziomie istotności $\alpha = 5\%$ sprawdzić hipotezę, że dbałość o ochronę środowiska przez przemysł istotnie wzrosła.

Rozwiązanie

Mamy porównać wskaźniki struktury p_1, p_2 z dwóch prób losowych, pobranych z tej samej populacji, ale w różnym czasie: obecnie i przed rokiem. Zatem hipoteza zerowa będzie miała postać (II.5.26):

$$H_0: p_1 = p_2.$$

Na podstawie treści zadania (chcemy sprawdzić, czy frakcja zakładów przemysłowych przestrzegających norm ekologicznych jest większa obecnie niż przed rokiem) stwierdzamy, iż hipoteza alternatywna musi być postaci:

$$H_1: p_1 < p_2,$$

przy czym p_1 oznacza frakcję w roku ubiegłym, p_2 w roku bieżącym.

Zmienna losowa zdefiniowana w zadaniu ma rozkład dwupunktowy: wybrany zakład przemysłowy przestrzega lub nie przestrzega norm ochrony środowiska. Ze względu na dużą liczebność analizowanych prób losowych zastosujemy model I do przeprowadzenia testu porównującego wskaźniki struktury.

Funkcja testowa przyjmuje następującą wartość:

$$Z = \frac{\frac{k_1}{n_1} - \frac{k_2}{n_2}}{\sqrt{\frac{\frac{k_1 + k_2}{n_1 + n_2} \cdot \left(1 - \frac{k_1 + k_2}{n_1 + n_2}\right)}{\frac{n_1 \cdot n_2}{n_1 + n_2}}}} = \frac{\frac{175}{250} - \frac{240}{300}}{\sqrt{\frac{\frac{175 + 240}{250 + 300} \cdot \left(1 - \frac{175 + 240}{250 + 300}\right)}{\frac{250 \cdot 300}{250 + 300}}}} = -2,71.$$

Z uwagi na postać hipotezy H_1 , mamy do czynienia z lewostronnym obszarem krytycznym, ograniczonym wartością krytyczną $-1,64$:

$$(-\infty, z(\alpha)] = (-\infty, -z(1 - \alpha)] = (-\infty; -1,64].$$

Ponieważ wartość statystyki należy do obszaru krytycznego – odrzucamy hipotezę zerową H_0 na korzyść hipotezy alternatywnej H_1 . Oznacza to, że w danym roku, w stosunku do roku poprzedniego, istotnie poprawiła się dbałość zakładów przemysłowych o środowisko naturalne. ■

II.5.1.7. Konsekwencje niespełnienia założenia o normalności rozkładu

Typy i wielkość błędów, jakie popełniane są, gdy założenie o normalności, niezbędne do zastosowania specyficznych testów nie jest spełnione przez analizowane dane, zostały empirycznie ocenione za pomocą tak zwanej metody Monte Carlo. Ogólny wniosek, jaki wypływa z tych badań, jest taki, że konsekwencje złamania tego założenia nie są na ogół tak poważne, jak sądzono wcześniej. Chociaż konkluzja ta nie powinna nikogo zachęcać do zaniechania troski o spełnienie założenia o normalności w jego badaniach, to jednak spowodowała ona niewątpliwie wzrost popularności stosowania testów statystycznych zależnych od typu rozkładu we wszystkich dziedzinach badań.

Mając na uwadze twierdzenie II.3.2, wynikające z centralnego twierdzenia granicznego, warto pamiętać, że dla większych prób testy wartości średnich są odporne na odstępstwa od normalności rozkładu, ale nie są odporne na wartości odstające. Natomiast testy wariancji są bezużyteczne przy istnieniu wyraźnych odstępstw od rozkładu normalnego. W szczególności test F na porównanie dwóch wariancji wykazuje brak odporności na niezachowanie normalności rozkładu. Wynika to stąd, że dla wyraźnie skośnych rozkładów – wariancja w ogóle nie jest dobrym miernikiem rozproszenia.

II.5.2. Testy nieparametryczne

Testy nieparametryczne służą do weryfikacji hipotezy zakładającej, że badana cecha ma rozkład prawdopodobieństwa o określonym kształcie, bez weryfikacji parametrów tego rozkładu. Na przykład chcemy potwierdzić (lub wykluczyć) hipotezę, że analizowana cecha wybranej populacji ma rozkład normalny.

Ze względu na wzajemnie jednoznaczną odpowiedniość pomiędzy funkcją dystrybuantą i funkcją gęstości (rozkład zmiennej ciągłej może być jednoznacznie zdefiniowany przez każdą z tych funkcji), dla zmiennej ciągłej hipoteza zerowa w testach nieparametrycznych może mieć jedną z następujących postaci:

$$H_0 : \text{dystrybuantą badanej cechy jest } F_0(x) \quad (\text{II.5.30})$$

lub:

$$H_0 : \text{gęstością prawdopodobieństwa badanej cechy jest } f_0(x) \quad (\text{II.5.31})$$

Natomiast dla zmiennej dyskretnej:

$$H_0 : \text{funkcją prawdopodobieństwa badanej cechy jest } P(x_i) = p_i \quad (\text{II.5.32})$$

O postaci F , f lub P wnioskujemy np. na podstawie histogramów dla wartości z próby losowej. Mogą to być funkcje całkowicie określone, albo też zależne od nieznanych parametrów.

Testy nieparametryczne opisane w tym rozdziale wymagają stosownego przygotowania empirycznych wartości zmiennej losowej, zgromadzonych w próbie. Dla pierwszego z nich, testu χ^2 -Pearsona (chi-kwadrat Pearsona) z wartości próby losowej tworzymy tzw. szereg rozdzielczy, którego konstrukcję omówiono w podrozdziale II.2.1. Natomiast do testu Kołmogorowa wartości w próbie porządkujemy niemalejąco.

II.5.2.1. Test χ^2 -Pearsona

Jako hipotezę zerową przyjmujemy wariant (II.5.30), zakładając, że dystrybuanta F_0 jest całkowicie określona. Szereg rozdzielczy, utworzony z doświadczalnych wartości zmiennej losowej x_1, x_2, \dots, x_n przedstawiony jest w tabeli II.5.1.

Tabela II.5.1

Szereg rozdzielczy utworzony z wyników doświadczenia losowego, zgromadzonych w próbie losowej

Nr klasy i	Granice klasy	Środek klasy	Liczność klasy n_i
1	(g_0, g_1)	\bar{x}_1	n_1
2	(g_1, g_2)	\bar{x}_2	n_2
...
k	(g_{k-1}, g_k)	\bar{x}_k	n_k
			$\sum_{i=1}^k n_i = n$

Jeżeli hipoteza H_0 jest prawdziwa, to prawdopodobieństwo p_i , że zmienna X przyjmie wartość należącą do i -tej klasy jest równe:

$$p_i = F_0(g_i) - F_0(g_{i-1}) \quad (\text{II.5.33})$$

Jeżeli zdarzenie (II.5.33) (wartość zmiennej należy do klasy i) zdefiniujemy jako „sukces”, to wartość teoretyczna liczby sukcesów w n niezależnych doświadczeniach, zgodnie z rozkładem Bernoulliego, wynosi: np_i . Wielkość np_i stanowi licznosc teoretyczną dla i -tej klasy, w odróżnieniu od licznosci doświadczalnej n_i .

Statystyka:

$$\chi_d^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (\text{II.5.34})$$

jest miarą różnic pomiędzy licznosciami doświadczalnymi i teoretycznymi. Dla ustalonej liczby klas k , przy prawdziwości hipotezy zerowej, dla $n \rightarrow \infty$, powyższa statystyka ma rozkład χ_d^2 dla $d = k - u$ stopni swobody, gdzie u – liczba wyznaczanych nieznaných parametrów.

Obszarem krytycznym testu dla poziomu istotności α jest przedział $[\chi^2(1-\alpha, k-u); +\infty)$. Zatem jeżeli $\chi_d^2 \in [\chi^2(1-\alpha, k-u); +\infty)$, to odrzucamy hipotezę zerową, czyli rozkładu badanej zmiennej losowej nie można uznać za zgodny z założonym rozkładem teoretycznym. Natomiast jeżeli $\chi_d^2 \notin [\chi^2(1-\alpha, k-u); +\infty)$, to nie ma podstaw do odrzucenia hipotezy zerowej o postaci rozkładu zmiennej losowej, dla którego dystrybuantą jest funkcja F_0 .

PRZYKŁAD II.5.5

Badając jednostkowy koszt produkcji pewnego artykułu produkowanego przez różne zakłady, wylosowano niezależnie 80 zakładów i otrzymano wyniki zebrane w tabeli II.5.2.

Tabela II.5.2

Wyniki oszacowania jednostkowego kosztu produkcji pogrupowane w klasy

Koszt jednostkowy [zł]	Liczba zakładów
20–40	10
40–60	16
60–80	24
80–100	18
100–120	12

Na poziomie istotności 10% zweryfikować hipotezę o normalności rozkładu jednostkowego kosztu produkcji tego artykułu.

Rozwiązanie

Kolejne etapy obliczeń, zamieszczono w tabeli II.5.3, a objaśniono pod tabelą. Na początku należy ustalić wartości zmiennej losowej, której rozkład mamy przetestować. Pier-

wotnie zebrane dane zostały już podzielone na klasy. Z każdej klasy wybieramy jej środek i na podstawie definicji II.2.2 dwie pierwsze kolumny tabeli II.5.3 stanowią szereg rozdzielczy badanej zmiennej losowej X .

Tabela II.5.3

Etapy obliczeń statystyki χ_d^2 do weryfikacji normalności rozkładu jednostkowego kosztu produkcji pewnego artykułu przez różne zakłady produkcyjne

\bar{x}_i	n_i	z_i	$\Phi(z_i)$	p_i	$n \cdot p_i$	$\frac{(n_i - np_i)^2}{np_i}$
30	10	-1,68	0,05	0,05	4	9,000
50	16	-0,87	0,19	0,14	11	2,273
70	24	-0,06	0,48	0,29	23	0,043
90	18	0,75	0,77	0,29	23	1,087
110	12	1,56	0,94	0,17	14	0,286
				0,06	5	5,000
				1,00	80	$\chi_d^2 = 17,689$

Hipoteza zerowa w tym teście będzie miała postać:

H_0 : dystrybucja zmiennej losowej X jest zgodna z dystrybucją rozkładu normalnego.

Natomiast hipoteza alternatywna H_1 jest jej zaprzeczeniem.

W ramach testu, przy założeniu normalności rozkładu zmiennej X , czyli prawdziwości H_0 , przeprowadzimy standaryzację tej zmiennej, a następnie wyznaczmy wartości dystrybucyjności zmiennej standaryzowanej Z , prawdopodobieństwa znalezienia się zmiennej w każdej klasie, a z nich – licznosci teoretyczne, które porównamy z licznosciami n_i , otrzymanymi w wyniku doświadczenia.

Chcąc przeprowadzić standaryzację zmiennej X , zgodnie ze wzorem z twierdzenia I.2.1, trzeba najpierw wyznaczyć przybliżone wartości średniej i odchylenia standardowego dla tej zmiennej, gdyż parametry te nie są znane z treści zadania. W tym celu posłużymy się estymatorami \hat{X} oraz \hat{S}^2 z tabeli II.4.3 dla obserwacji różnorodnych. Jeżeli potraktujemy licznosci n_i jako wagi p_i , to estymator średniej wynosi:

$$\hat{\mu} = \hat{X} = \frac{\sum_{i=1}^5 n_i \cdot \bar{x}_i}{\sum_{i=1}^5 n_i} = \frac{10 \cdot 30 + 16 \cdot 50 + 24 \cdot 70 + 18 \cdot 90 + 12 \cdot 110}{10 + 16 + 24 + 18 + 12} = \frac{5720}{80} = 71,5,$$

a estymator odchylenia standardowego wynosi:

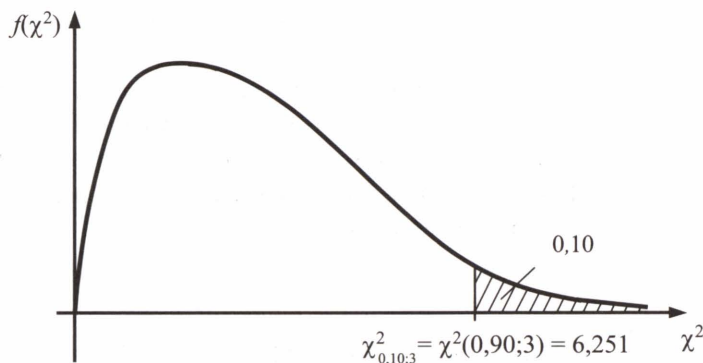
$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^5 n_i \cdot (\bar{x}_i - \hat{X})^2}{\sum_{i=1}^5 n_i}} = \sqrt{\frac{10 \cdot (30 - 71,5)^2 + 16 \cdot (50 - 71,5)^2 + 24 \cdot (70 - 71,5)^2 + 18 \cdot (90 - 71,5)^2 + 12 \cdot (110 - 71,5)^2}{80}} = 24,6526.$$

Standaryzowane wartości zmiennej losowej wyznaczamy zgodnie ze wzorem:

$$Z = \frac{X - \mu}{\sigma}.$$

Następnie z tablic dystrybuanty standaryzowanego rozkładu normalnego $N(0,1)$ odczytujemy wartości tej funkcji $\Phi(Z_i)$ dla wyliczonych Z_i , przy czym, dla wartości ujemnych Z_i stosujemy przekształcenie: $\Phi(Z_i) = 1 - \Phi(-Z_i)$, np.: $\Phi(-1,68) = 1 - \Phi(1,68) = 1 - 0,953521 = 0,046479$. Na podstawie wartości dystrybuanty wyznaczamy teoretyczne prawdopodobieństwa zdarzeń, że koszt produkcji znajdzie się w określonej klasie. Korzystamy przy tym z definicji dystrybuanty I.2.2, z której wynika: $P(Z \in (z_1, z_2)) = \Phi(z_2) - \Phi(z_1)$.

Mając w pełni określoną przestrzeń probabilistyczną, czyli zakres i przedziały dla wartości zmiennej losowej oraz prawdopodobieństwa, że przyjmie ona wartość w określonym przedziale – z łatwością wyznaczamy teoretyczne licznosci wszystkich przedziałów $n \cdot p_i$.



Rys. II.5.6. Prawostronny obszar krytyczny w teście weryfikującym normalność rozkładu jednostkowego kosztu produkcji

Na koniec pozostaje zastosowanie wzoru (II.5.34) do wyznaczenia wartości funkcji testowej, którą porównujemy z wartością krytyczną testu dla zadanego poziomu istotności 0,10 oraz liczby stopni swobody $6-3=3$, gdzie 6 to liczba klas, a 3 to liczba wyznaczanych parametrów $\hat{X}, \hat{\sigma}, \chi_d^2$. Odczytana z tablic rozkładu chi-kwadrat wartość krytyczna to $\chi_{0,10;3}^2 = 6,251$. Jak widać $\chi_d^2 \in [\chi^2(0,90;3); +\infty)$ (rys. II.5.6), zatem hipotezę o normalności rozkładu należy odrzucić. Rozkład jednostkowego kosztu produkcji określonego artykułu jest inny niż normalny. ■

PRZYKŁAD II.5.6

W doświadczeniu mającym na celu sprawdzenie, czy konkretna kostka do gry jest symetryczna wykonano 100 rzutów tą kostką, otrzymując następujące liczebności wyników (tab. II.5.4).

Tabela II.5.4
Szereg rozdzielczy wyników 100 rzutów kostką do gry

x_i	1	2	3	4	5	6
n_i	20	18	14	15	10	23

Rozwiązanie

Zadanie sprowadza się do wykonania nieparametrycznego testu statystycznego, weryfikującego hipotezę o jednostajności rozkładu liczby uzyskanych oczek podczas rzutu kostką. Innymi słowy: chcemy odpowiedzieć na pytanie czy prawdopodobieństwo otrzymania dowolnej liczby oczek jest stałe i wynosi 1/6. Zatem hipoteza zerowa ma postać:

H_0 : rozkład liczby uzyskiwanych oczek w rzucie kostką jest jednostajny (tab. II.5.5).

Tabela II.5.5
Teoretyczny rozkład prawdopodobieństwa wyników rzutu kostką do gry

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

Hipoteza alternatywna może w tej sytuacji brzmieć:

H_1 : rozkład liczby oczek uzyskiwany przy rzucie kostką jest inny niż jednostajny.

Przy prawdziwości hipotezy zerowej licznosci dla każdej liczby oczek są jednakowe i wynoszą:

$$n \cdot p_i = \frac{100}{6} = 16,67.$$

Są to nasze licznosci teoretyczne. Zgodnie z postacia statystyki w tym teście (II.5.34), wyliczamy jej wartosc:

$$\begin{aligned}\chi_d^2 &= \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \\ &= \frac{(20 - 16,67)^2 + (18 - 16,67)^2 + (14 - 16,67)^2 + (15 - 16,67)^2 + (10 - 16,67)^2 + (23 - 16,67)^2}{16,67} = \\ &= \frac{11,0889 + 1,7689 + 7,1289 + 2,7889 + 44,4889 + 40,0689}{16,67} = 6,44.\end{aligned}$$

Przy poziomie istotnosci $\alpha = 0,05$, odczytana z tablic rozkladu chi-kwadrat wartosc krytyczna dla (liczba klas $- 1$) = 5 stopni swobody wynosi $\chi_{0,05;5}^2 = \chi^2(0,95; 5) = 11,07$. Otrzymana wartosc statystyki $\chi_d^2 = 6,44$ nie wpada do obszaru krytycznego okreslonego przez przedzial $[\chi^2(1 - \alpha, k - 1); +\infty) = [11,07; +\infty)$. Zatem nie ma podstaw do odrzucenia hipotezy zerowej mowiacej o jednostajnosci badanego rozkladu.

■

II.5.2.2. Test Kolmogorowa

Weryfikujac za pomoca testu Kolmogorowa hipoteze H_0 , ze cecha X , jako ciagla zmienna losowa, ma dystrybuante F_0 o znanych parametrach, ukladamy wartosci z proby losowej w niemalejacy ciag: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Dla modelu, w ktorym zmienna losowa bedzie tak uporzadkowana, jako statystyke testowa Kolmogorow przyjal nastepujaca funkcje:

$$D_n = \sup_x |F_0(x) - F_E(x)| \quad (\text{II.5.35})$$

gdzie $F_E(x)$ – dystrybuanta empiryczna z wynikow proby, uporzadkowanych w kolejnosci niemalejacej, zdefiniowana nastepujaco:

$$F_E(x) = \begin{cases} 0 & \text{dla } x < x_1 \\ \frac{i}{n} & \text{dla } x_{(i)} \leq x < x_{(i+1)} \quad 1 \leq i \leq n-1 \\ 1 & \text{dla } x \geq x_{(n)} \end{cases} \quad (\text{II.5.36})$$

Statystyka D_n – w przypadku prawdziwosci hipotezy zerowej H_0 – ma rozklad niezalezny od niej. Na podstawie tego rozkladu, dla poziomu istotnosci α , sporzadzone zostaly tablice kwantyli $d_n(1 - \alpha)$ spelniajace rownosc:

$$P(D_n \geq d_n(1 - \alpha)) = \alpha \quad (\text{II.5.37})$$

W praktyce, przy stosowaniu tego testu, postępujemy zgodnie z następującymi etapami:

1. Po uporządkowaniu wyników pomiaru w kolejności niemalejącej, obliczamy wszystkie różnice:

$$\frac{i}{n} - F_0(x_{(i)}) \quad \text{oraz} \quad F_0(x_{(i)}) - \frac{i-1}{n} \quad \text{dla } i = 1, \dots, n \quad (\text{II.5.38})$$

2. Wybieramy największą z ich wartości bezwzględnych:

$$d_n = \max_i \left(\left| \frac{i}{n} - F_0(x_{(i)}) \right|, \left| F_0(x_{(i)}) - \frac{i-1}{n} \right| \right) \quad (\text{II.5.39})$$

3. Dla zadanego poziomu istotności α , odczytujemy z tablicy 10 (dodatek) kwantyle pełniący rolę wartości krytycznej $d_n(1-\alpha)$, która wyznacza obszar krytyczny w postaci następującego przedziału:

$$[d_n(1-\alpha); 1] \quad (\text{II.5.40})$$

4. Wnioskowanie statystyczne przebiega następująco:

- 4a. jeżeli $d_n \in [d_n(1-\alpha); 1]$, to odrzucamy hipotezę zerową, czyli rozkładu badanej zmiennej losowej nie można uznać za zgodny z założonym rozkładem teoretycznym;
- 4b. jeżeli $d_n \notin [d_n(1-\alpha); 1]$, to nie ma podstaw do odrzucenia hipotezy zerowej o postaci rozkładu zmiennej losowej, dla którego dystrybuantą jest funkcja F_0 .

Uwaga. Test Kołmogorowa, w porównaniu z testem χ^2 -Pearsona, wymaga mniej licznej próby, przy zachowaniu tej samej mocy testu. Jest to zatem test znacznie czulszy.

PRZYKŁAD II.5.7

W tabeli II.5.6 zamieszczono wykaz średnich dobowych stężeń tlenków azotu NO_x podanych w $\mu\text{g}/\text{m}^3$, powstających podczas spalania paliw w kotłach energetycznych na pewnym terenie przemysłowym. Pomiary wykonywano codziennie przez drugi kwartał roku (92 dni). W tabeli zmienna T oznacza kolejny dzień pomiarowy. Sprawdzić normalność rozkładu uzyskanych wyników na poziomie istotności 0,01.

Tabela II.5.6
Średnie dobowe stężenia tlenków azotu

T	NO_x	T	NO_x	T	NO_x	T	NO_x
1	31,6	24	7,2	47	5,5	70	7,2
2	27,0	25	11,2	48	6,6	71	9,2
3	4,1	26	12,0	49	9,9	72	14,0
4	26,6	27	23,4	50	0,1	73	5,5

Tabela II.5.6 cd.

T	NO_x	T	NO_x	T	NO_x	T	NO_x
5	15,9	28	12,5	51	24,7	74	18,0
6	0,1	29	9,1	52	10,8	75	12,4
7	10,9	30	9,6	53	14,9	76	44,4
8	47,9	31	9,9	54	16,2	77	11,3
9	13,0	32	7,8	55	12,4	78	11,1
10	17,0	33	9,4	56	12,3	79	15,8
11	10,3	34	47,8	57	13,0	80	12,3
12	28,5	35	20,0	58	47,9	81	16,4
13	9,3	36	9,1	59	10,9	82	35,3
14	9,2	37	4,7	60	13,5	83	31,8
15	12,3	38	12,0	61	10,9	84	11,4
16	12,0	39	11,8	62	0,0	85	12,2
17	10,9	40	12,0	63	13,7	86	13,2
18	22,5	41	15,1	64	6,6	87	4,7
19	14,5	42	13,5	65	12,9	88	6,6
20	44,4	43	15,1	66	8,4	89	13,1
21	9,9	44	12,6	67	12,5	90	30,4
22	47,9	45	9,9	68	6,6	91	12,3
23	9,4	46	0,0	69	11,0	92	8,6

Rozwiązanie

W celu przetestowania normalności wyników pomiaru zanieczyszczenia powietrza tlenkami azotu, zostaną przeprowadzone dwa nieparametryczne testy istotności: test chi-kwadrat i test Kołmogorowa.

Hipoteza zerowa w teście ma postać:

H_0 : dystrybuanta zmiennej losowej X jest zgodna z dystrybuantą rozkładu normalnego

Na potrzeby testu chi-kwadrat, z zebranych pomiarów utworzono szereg rozdzielczy (tab. II.5.7), w którym pierwotnie przyjęto liczbę klas 10, wyznaczoną zgodnie z sugestią zamieszczoną w podrozdziale II.2.1, jako pierwiastek z liczby danych pomiarowych $\sqrt{92} \cong 10$. Jednakże liczności doświadczalne i teoretyczne w klasach 8 i 9 były bardzo małe,

zatem połączono te klasy w jedną razem z klasą numer 10 i wykonano obliczenia dla 8 klas. Są one analogiczne do obliczeń w przykładzie II.5.5. Przez \bar{x}_i oznaczono środek klasy i .

W ramach testu, przy założeniu normalności rozkładu zmiennej X , czyli prawdziwości H_0 , przeprowadzamy standaryzację tej zmiennej, a następnie wyznaczamy wartości dystrybucyjności zmiennej standaryzowanej Z , prawdopodobieństwa znalezienia się zmiennej w każdej klasie, a z nich licznosci teoretyczne, które porównujemy z licznosciami doświadczalnymi n_i .

Chcąc przeprowadzić standaryzację zmiennej X (po podziale jej wartości na klasy), zgodnie ze wzorem z twierdzenia I.2.1, najpierw wyznaczamy przybliżone wartości średniej i odchylenia standardowego dla tej zmiennej, z próby 8-elementowej. W tym celu posłużymy się estymatorami \hat{X} oraz \hat{S}^2 z tabeli II.4.3 dla obserwacji różnorodokładnych. Jeżeli potraktujemy licznosci n_i jako wagi p_i , to estymator średniej wynosi:

$$\hat{\mu} = \hat{X} = \frac{\sum_{i=1}^8 n_i \cdot \bar{x}_i}{\sum_{i=1}^8 n_i} = \frac{1310,4}{92} \cong 14,24 \mu\text{g/m}^3,$$

a estymator odchylenia standardowego wynosi:

$$\hat{\sigma} = \hat{S} = \sqrt{\frac{\sum_{i=1}^8 n_i \cdot (\bar{x}_i - \hat{X})^2}{\sum_{i=1}^8 n_i}} = 9,3476 \mu\text{g/m}^3.$$

Tabela II.5.7

Etapy obliczeń statystyki χ_d^2 do weryfikacji normalności rozkładu średnich dobowych stężeń tlenków azotu

Lp.	NO _x [μg/m ³]	n_i	\bar{x}_i	$z_i = \frac{\bar{x}_i - \hat{X}}{\hat{\sigma}}$	$\Phi(z_i)$	p_i	$n \cdot p_i$	$\frac{(n_i - np_i)^2}{np_i}$
1	0,0÷4,8	7	2,4	-1,27	0,10	0,10	9,2	0,526
2	4,8÷9,6	19	7,2	-0,75	0,23	0,13	12,0	4,083
3	9,6÷14,4	40	12,0	-0,24	0,41	0,18	16,6	32,985
4	14,4÷19,2	10	16,8	0,27	0,61	0,20	18,4	3,835
5	19,2÷24,0	3	21,6	0,79	0,78	0,17	15,6	10,177
6	24,0÷28,8	4	26,4	1,30	0,90	0,12	11,0	4,454
7	28,8÷33,6	3	31,2	1,81	0,96	0,06	5,5	1,136
8	> 33,6	6	40,8	2,84	1,00	0,04	3,7	1,430
Suma:		92	Suma:			1,00	92	$\chi_d^2 = 58,626$

Reszta obliczeń (tab. II.5.7) przebiega analogicznie do obliczeń w przykładzie II.5.5. Na końcu, według wzoru (II.5.34), wyznaczamy wartość funkcji testowej, którą porównujemy z wartością krytyczną testu dla zadanego poziomu istotności 0,01 oraz liczby stopni swobody $8-3=5$, gdzie 8 to liczba klas, a 3 to liczba wyznaczanych parametrów $\hat{X}, \hat{\sigma}, \chi_d^2$. Odczytana z tablic rozkładu chi-kwadrat wartość krytyczna to $\chi_{0,01;5}^2 = \chi^2(0,99; 5) = 15,086$. Jak widać, $\chi_d^2 \in [\chi_{0,01;5}^2; +\infty)$, zatem hipotezę o normalności rozkładu należy odrzucić. Test chi-kwadrat wykazał, że średnie dobowe stężenia tlenków azotu mają rozkład inny niż normalny.

Na potrzeby testu Kołmogorowa wartości próby losowej NO_x zostały ułożone w ciąg niemalejący (w tabeli II.5.8 są to kolumny $x_{(i)}$). Następnie wyznaczono wartości dystrybuanty empirycznej według wzoru (II.5.36) oraz wartości dystrybuanty teoretycznej za pomocą tablicy 1 (dodatek), po wykonaniu standaryzacji zmiennej X :

$$A \stackrel{\text{ozn.}}{=} F_0(x_{(i)}) = \Phi_0\left(\frac{x_{(i)} - \hat{x}}{\hat{\sigma}}\right) = \Phi_0\left(\frac{x_{(i)} - 14,5}{10,22621}\right),$$

przy czym wartości estymatorów nieobciążonych średniej \hat{x} i odchylenia standardowego $\hat{\sigma}$ wyznaczono z próby 92-elementowej, na podstawie wzorów w tabeli II.4.3 dla obserwacji jednakowo dokładnych.

Kolejne etapy obliczeń – zamieszczone w tabeli II.5.8 – to obliczenie różnic (II.5.38), które w tabeli oznaczono B i C :

$$B \stackrel{\text{ozn.}}{=} \frac{i}{92} - F_0(x_{(i)})$$

oraz:

$$C \stackrel{\text{ozn.}}{=} F_0(x_{(i)}) - \frac{i-1}{92} \quad \text{dla} \quad i=1, 2, \dots, 92.$$

Wyznaczona według wzoru (II.5.39) wartość funkcji testowej wynosi:

$$d_{92} = \max_i \left(\left| \frac{i}{92} - F_0(x_{(i)}) \right|, \left| F_0(x_{(i)}) - \frac{i-1}{92} \right| \right) = \max_i (0,238; 0,227) = 0,238.$$

Na podstawie tablicy 10 (dodatek) widzimy, że wraz ze wzrostem liczności próby losowej n , maleje wartość krytyczna dla zadanego poziomu istotności. Odczytany kwantyl statystyki D_n Kołmogorowa, dla liczności próby równej $n = 92$, pełniący rolę wartości krytycznej wynosi $d_{0,01; 92} = d_{92}(0,99) = 0,168$. Widzimy więc, że obliczona wartość statystyki należy do obszaru krytycznego testu $d = 0,238 \in [d_{92}(1 - 0,01); 1]$. Stąd wniosek podobny jak w teście chi-kwadrat, że rozkładu zmiennej losowej opisującej średnie dobowe stężenia tlenków azotu nie można uznać za zgodny z rozkładem normalnym.

Na rysunku II.5.7 zamieszczono histogram rozkładu średnich dobowych stężeń tlenków azotu z nałożoną krzywą funkcji gęstości rozkładu normalnego. Widać, że rzeczywiście rozkład empiryczny znacznie odbiega od założonego rozkładu normalnego.

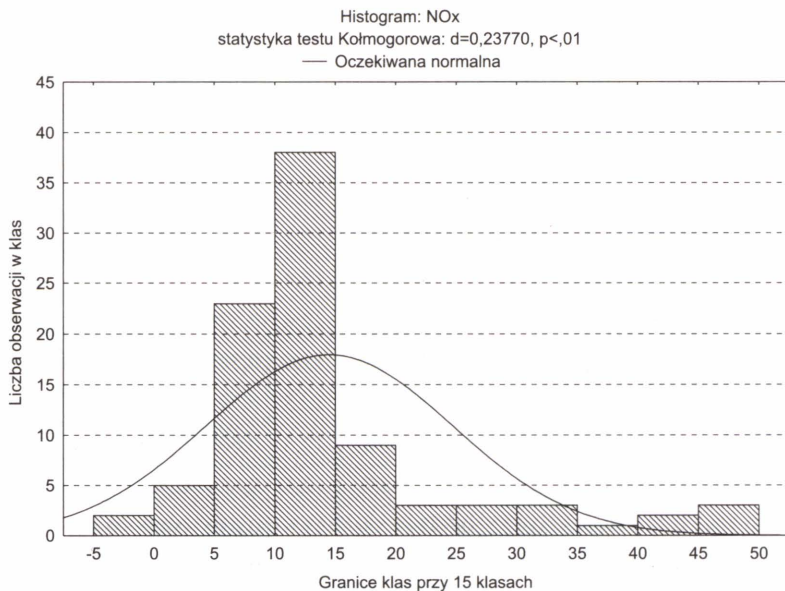
Tabela II.5.8

Etapy obliczeń statystyki d do weryfikacji normalności rozkładu średnich dobowych stężeń tlenków azotu

Lp.	$x_{(i)}$	$\frac{i}{n}$	A	$\frac{i-1}{n}$	B	C	Lp.	$x_{(i)}$	$\frac{i}{n}$	A	$\frac{i-1}{n}$	B	C
1	0,0	0,011	0,078	0,000	0,067	0,078	31	10,3	0,337	0,341	0,326	0,004	0,015
2	0,0	0,022	0,078	0,011	0,056	0,067	32	10,8	0,348	0,359	0,337	0,011	0,022
3	0,1	0,033	0,080	0,022	0,047	0,058	33	10,9	0,359	0,362	0,348	0,004	0,015
4	0,1	0,043	0,080	0,033	0,036	0,047	34	10,9	0,370	0,362	0,359	0,007	0,004
5	4,1	0,054	0,155	0,043	0,100	0,111	35	10,9	0,380	0,362	0,370	0,018	0,007
6	4,7	0,065	0,169	0,054	0,104	0,115	36	10,9	0,391	0,362	0,380	0,029	0,018
7	4,7	0,076	0,169	0,065	0,093	0,104	37	11,0	0,402	0,366	0,391	0,036	0,025
8	5,5	0,087	0,189	0,076	0,102	0,113	38	11,1	0,413	0,370	0,402	0,043	0,032
9	5,5	0,098	0,189	0,087	0,092	0,102	39	11,2	0,424	0,373	0,413	0,050	0,040
10	6,6	0,109	0,220	0,098	0,111	0,122	40	11,3	0,435	0,377	0,424	0,058	0,047
11	6,6	0,120	0,220	0,109	0,100	0,111	41	11,4	0,446	0,381	0,435	0,065	0,054
12	6,6	0,130	0,220	0,120	0,089	0,100	42	11,8	0,457	0,396	0,446	0,061	0,050
13	6,6	0,141	0,220	0,130	0,079	0,089	43	12,0	0,467	0,403	0,457	0,064	0,053
14	7,2	0,152	0,238	0,141	0,085	0,096	44	12,0	0,478	0,403	0,467	0,075	0,064
15	7,2	0,163	0,238	0,152	0,075	0,085	45	12,0	0,489	0,403	0,478	0,086	0,075
16	7,8	0,174	0,256	0,163	0,082	0,093	46	12,0	0,500	0,403	0,489	0,097	0,086
17	8,4	0,185	0,275	0,174	0,091	0,102	47	12,2	0,511	0,411	0,500	0,100	0,089
18	8,6	0,196	0,282	0,185	0,086	0,097	48	12,3	0,522	0,415	0,511	0,107	0,096
19	9,1	0,207	0,299	0,196	0,092	0,103	49	12,3	0,533	0,415	0,522	0,118	0,107
20	9,1	0,217	0,299	0,207	0,081	0,092	50	12,3	0,543	0,415	0,533	0,129	0,118
21	9,2	0,228	0,302	0,217	0,074	0,085	51	12,3	0,554	0,415	0,543	0,140	0,129
22	9,2	0,239	0,302	0,228	0,063	0,074	52	12,4	0,565	0,419	0,554	0,147	0,136
23	9,3	0,250	0,306	0,239	0,056	0,066	53	12,4	0,576	0,419	0,565	0,157	0,147
24	9,4	0,261	0,309	0,250	0,048	0,059	54	12,5	0,587	0,422	0,576	0,164	0,154
25	9,4	0,272	0,309	0,261	0,037	0,048	55	12,5	0,598	0,422	0,587	0,175	0,164
26	9,6	0,283	0,316	0,272	0,033	0,044	56	12,5	0,609	0,422	0,598	0,186	0,175
27	9,9	0,293	0,326	0,283	0,033	0,044	57	12,6	0,620	0,426	0,609	0,193	0,182
28	9,9	0,304	0,326	0,293	0,022	0,033	58	12,9	0,630	0,438	0,620	0,193	0,182
29	9,9	0,315	0,326	0,304	0,011	0,022	59	13,0	0,641	0,442	0,630	0,200	0,189
30	9,9	0,326	0,326	0,315	0,000	0,011	60	13,0	0,652	0,442	0,641	0,210	0,200

Tabela II.5.8 cd.

Lp.	$x_{(i)}$	$\frac{i}{n}$	A	$\frac{i-1}{n}$	B	C	Lp.	$x_{(i)}$	$\frac{i}{n}$	A	$\frac{i-1}{n}$	B	C
61	13,1	0,663	0,446	0,652	0,217	0,207	77	20,0	0,837	0,705	0,826	0,132	0,121
62	13,2	0,674	0,449	0,663	0,224	0,214	78	22,5	0,848	0,783	0,837	0,065	0,054
63	13,5	0,685	0,461	0,674	0,224	0,213	79	23,4	0,859	0,808	0,848	0,051	0,040
64	13,5	0,696	0,461	0,685	0,235	0,224	80	24,7	0,870	0,841	0,859	0,029	0,018
65	13,7	0,707	0,469	0,696	0,238	0,227	81	26,6	0,880	0,882	0,870	0,001	0,012
66	14,0	0,717	0,481	0,707	0,237	0,226	82	27,0	0,891	0,889	0,880	0,002	0,009
67	14,5	0,728	0,500	0,717	0,228	0,217	83	28,5	0,902	0,915	0,891	0,012	0,023
68	14,9	0,739	0,516	0,728	0,224	0,213	84	30,4	0,913	0,940	0,902	0,027	0,038
69	15,1	0,750	0,523	0,739	0,227	0,216	85	31,6	0,924	0,953	0,913	0,029	0,040
70	15,1	0,761	0,523	0,750	0,237	0,227	86	31,8	0,935	0,955	0,924	0,020	0,031
71	15,8	0,772	0,551	0,761	0,221	0,210	87	35,3	0,946	0,979	0,935	0,033	0,044
72	15,9	0,783	0,554	0,772	0,228	0,217	88	44,4	0,957	0,998	0,946	0,042	0,053
73	16,2	0,793	0,566	0,783	0,227	0,217	89	44,4	0,967	0,998	0,957	0,031	0,042
74	16,4	0,804	0,574	0,793	0,231	0,220	90	47,8	0,978	0,999	0,967	0,021	0,032
75	17,0	0,815	0,597	0,804	0,219	0,208	91	47,9	0,989	0,999	0,978	0,010	0,021
76	18,0	0,826	0,634	0,815	0,192	0,181	92	47,9	1,000	0,999	0,989	0,001	0,010



Rys. II.5.7. Histogram średnich dobowych stężeń tlenków azotu z nałożoną krzywą Gaussa

II.6. Regresja

Pojęcie **regresji** wiąże się ściśle z modelowaniem matematycznym analizowanego zjawiska przyrodniczego, ekonomicznego, socjologicznego czy technicznego itp. Chodzi bowiem o matematyczny opis zależności między różnymi zmiennymi losowymi, aby następnie wykorzystać go do tzw. prognozowania, a czasem do uzupełniania brakujących danych pomiarowych. Znajomość równań regresji to znajomość zależności między zmiennymi losowymi, wyrażona za pomocą funkcji opisującej liczbowo ten związek. Wzór funkcyjny jest bardzo wygodny do wyznaczenia wartości jednej zmiennej przy założeniu konkretnej wartości drugiej zmiennej. Jednakże prognozowanie takie jest najczęściej obarczone pewnym błędem.

Oczywiście oszacowanie będzie tym lepsze, im suma bezwzględnych wartości błędów, jakie popełniamy, zastępując rzeczywiste wartości zmiennej losowej – wartościami otrzymanymi z równania regresji, będzie mniejsza. W interpretacji geometrycznej oznacza to, że oszacowanie jest tym lepsze, im na wykresie rozrzutu punkty bardziej skupiają się wokół linii regresji. Mówiąc „linia regresji”, nie mamy na myśli jedynie wykresu funkcji liniowej, ale wykresy dowolnych funkcji regresji, również nieliniowych:

$$Y = g(X) \quad (\text{II.6.1})$$

Jako miary rozproszenia punktów wokół linii regresji używa się zwykle tzw. **wariancji resztowej**, określonej wzorem:

$$V(Z) = E[(Z - E(Z))^2] \quad (\text{II.6.2})$$

przy czym zmienna Z , określona następująco: $z_i = y_i - g(x_i)$, to tzw. **błąd losowy** (różnica pomiędzy rzeczywistą wartością zmiennej Y i otrzymaną z równania regresji (II.6.1). Dla takiego określenia błędu losowego, otrzymujemy, że $E(Z) = 0$. Zatem wariancja $V(Z)$ redukuje się do wyrażenia:

$$V(Z) = E[(Z - E(Z))^2] = E(Z^2) = E[(Y - g(X))^2] \quad (\text{II.6.3})$$

gdzie:

- y – zaobserwowana wartość zmiennej zależnej,
- $g(x)$ – prognozowana wartość zmiennej zależnej, wynikająca z przyjętego modelu.

Pierwiastek kwadratowy z wariancji resztowej nazywamy **standardowym błędem oceny** lub **średnim błędem resztowym** i oznaczamy symbolem $\sigma(Z)$.

Jak wiadomo, gdy zmienna losowa (X, Y) ma rozkład normalny, regresja Y względem X oraz X względem Y jest funkcją liniową (por. podrozdział I.3.6).

W statystyce matematycznej nie ma analitycznych sposobów, które by umożliwiały optymalny wybór odpowiedniej postaci funkcji regresji. Miarę zaufania do modelu regresji (II.6.1), opisującego zmienność tzw. **zmiennej zależnej (objaśnianej)** względem **zmiennej niezależnej (objaśniającej)**, stanowi kwadrat współczynnika korelacji krzywoliniowej q liczony następująco:

$$q^2 = 1 - \frac{\sum_{i=1}^n [y_i - g(x_i)]^2}{\sum_{i=1}^n [y_i - \hat{y}]^2} \quad (\text{II.6.4})$$

gdzie:

- y_i – zaobserwowane wartości zmiennej zależnej,
- \hat{y} – wartość przeciętna zmiennej zależnej,
- $g(x_i)$ – prognozowane wartości zmiennej zależnej, wynikające z przyjętego nieliniowego modelu.

Wartość tego współczynnika może stanowić kryterium doboru odpowiedniej postaci funkcji g . Im q^2 jest bliższe 1, tym funkcja g jest lepiej dopasowana do danych empirycznych.

Po przeanalizowaniu wykresu rozrzutu utworzonego z empirycznych wartości zmiennych: zależnej i niezależnej, często obserwuje się na tyle duże nieregularności, że trudno jest sformułować globalne reguły rządzące związkiem między analizowanymi zmiennymi. W tym wypadku oznacza to, że trudno jest znaleźć jeden typ funkcji, który by najlepiej opisywał badaną zależność.

W praktyce często zdarza się sytuacja, w której zmienna objaśniana zależy od wielu zmiennych objaśniających jednocześnie. Wówczas badane zjawisko ma charakter wielowymiarowy, a do jego opisu szukamy postaci funkcji uwzględniającej wpływ wielu czynników na jedną zmienną. Jest to tzw. **regresja wielowymiarowa**:

$$Y = F(X, a) \quad (\text{II.6.5})$$

gdzie:

- Y – zmienna losowa zależna (objaśniana),
- X – zmienna wielowymiarowa odpowiadająca czynnikom wpływającym na Y , stanowiących zmienne niezależne (objaśniające),
- a – wektor parametrów modelu.

Najprostszą formą regresji wielowymiarowej jest tzw. **liniowa regresja wieloraka**:

$$Y = F(X, a) = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_m \cdot x_m \quad (\text{II.6.6})$$

Model liniowej regresji wielorakiej zakłada liniowy wpływ każdej ze zmiennych objaśniających na zmienną objaśnianą lub przynajmniej zachowuje liniowość względem parametrów:

$$Y = F(X, a) = a_0 + a_1 \cdot g_1(x_1) + a_2 \cdot g_2(x_2) + \dots + a_m \cdot g_m(x_m) \quad (\text{II.6.7})$$

Estymację parametrów liniowej regresji wielorakiej wykonuje się metodą najmniej szych kwadratów (por. przykład II.4.1 w podrozdz. II.4.2.1).

W bardziej ogólnym przypadku poszczególne składniki sumy tworzące funkcję F (II.6.5), nie muszą mieć charakteru liniowego ani względem składowych zmiennej X , ani względem parametrów, lecz są indywidualnie dobranymi funkcjami wynikającymi z badania zależności zmiennej objaśnianej w powiązaniu z każdą ze zmiennych objaśniających osobno. Z tak

wybranych postaci funkcji g tworzony jest model wielowymiarowy, czyli globalna funkcja nieliniowa F :

$$Y = F(X, a) = a_0 + g_1(x_1) + g_2(x_2) + \dots + g_m(x_m) \quad (\text{II.6.8})$$

W jeszcze większym uogólnieniu, poszczególne funkcje g mogą być – w obrębie jednej funkcji F – powiązane ze sobą zależnościami inną niż addytywna. Do zależności addytywnej odpowiednie są te czynniki kształtujące zmienną zależną, dla których we wstępnej analizie wykazano lub wiadomo z charakteru zmiennych, że na związek między danym czynnikiem i zmienną zależną nie mają istotnego wpływu inne zmienne objaśniające.

Odpowiednikiem współczynnika korelacji krzywoliniowej q w przypadku wielowymiarowym jest **wielowymiarowy współczynnik korelacji krzywoliniowej**. Jego definicja jest analogiczna do (II.6.4), dlatego zachowano to samo oznaczenie q .

Jeżeli $Y = F(X_1, X_2, \dots, X_m, a)$ jest hiperpowierzchnią drugiego rodzaju wyznaczoną na podstawie próbki (X, Y) , gdzie $X = (X_1, X_2, \dots, X_m)$ to m -wymiarowy wektor zmiennych niezależnych (objaśniających), $a = (a_1, a_2, \dots, a_k)$ to wektor parametrów funkcji F , to kwadrat współczynnika korelacji krzywoliniowej q^2 zdefiniowany jest następująco:

$$q^2 = 1 - \frac{\sum_{i=1}^n [y_i - F_i(X, a)]^2}{\sum_{i=1}^n [y_i - E(y)]^2} = \frac{\sum_{i=1}^n [F_i(X, a) - E(y)]^2}{\sum_{i=1}^n [y_i - E(y)]^2} \Leftrightarrow \quad (\text{II.6.9})$$

$$\Leftrightarrow q^2 = 1 - \frac{NSK}{CSK} = \frac{WSK}{CSK}$$

gdzie:

n – liczność próby losowej, na podstawie której estymujemy parametry modelu regresji;

$E(y)$ – wartość oczekiwana zaobserwowanych wartości zmiennej zależnej;

$\sum_{i=1}^n [y_i - E(y)]^2$ – całkowite rozproszenie zmiennej zależnej względem jej wartości przeciętnej (CSK);

$\sum_{i=1}^n [y_i - F_i(X, a)]^2$ – suma kwadratów odchylek pomiędzy wartościami zmiennej zależnej z próby y_i a jej wartościami modelowymi, wyrażająca część niewyjaśnioną modelem regresji nieliniowej (NSK);

$\sum_{i=1}^n [F_i(X, a) - E(y)]^2$ – suma kwadratów odchylek pomiędzy wartościami modelowymi zmiennej zależnej a wartością przeciętną z zaobserwowanych wartości tej zmiennej, wyrażająca część wyjaśnioną modelem regresji nieliniowej (WSK).

Współczynnik ten określa zgodność wartości modelowych z empirycznymi i może być interpretowany jako udział wariancji części wyjaśnionej przez model regresji nieliniowej w całkowitym rozproszeniu zmiennej zależnej względem jej wartości przeciętnej. W przy-

padku modelu liniowej regresji wielorakiej (II.6.6) nosi on nazwę współczynnika determinacji i jest oznaczany R^2 (wzór II.6.10). Jego wartość może być wyznaczona również na podstawie macierzy korelacyjnej K , zawierającej współczynniki korelacji liniowej r_{ij} (zdefiniowane w podrozdziale I.3.5), wyznaczone dla wszystkich par zmiennych występujących w modelu liniowej regresji wielorakiej.

$$R^2 = 1 - \frac{\det K}{\det K_0} \quad (\text{II.6.10})$$

gdzie:

$$K = \begin{pmatrix} 1 & r_{01} & r_{02} & \dots & r_{0m} \\ r_{10} & 1 & r_{12} & \dots & r_{1m} \\ r_{20} & r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m0} & r_{m1} & r_{m2} & \dots & 1 \end{pmatrix} \quad (\text{II.6.11})$$

K – macierz korelacyjna, zawierająca współczynniki korelacji zupełnej pomiędzy wszystkimi parami zmiennych (r_{ij});

$\det K$ – wyznacznik macierzy korelacyjnej;

$\det K_0$ – wyznacznik podmacierzy, która powstaje ze skreślenia pierwszego wiersza i pierwszej kolumny w macierzy K , czyli współczynników korelacji dotyczących zmiennej zależnej (objaśnianej modelem regresji).

Wielkość $1-R^2$ stanowi współczynnik niezgodności modelu z empirycznymi wartościami zmiennej objaśnianej, które są wykorzystywane w estymacji parametrów modelu regresji wielorakiej.

Na podstawie wartości q^2 lub R^2 można zweryfikować hipotezę o równości wariancji części wyjaśnionej przez model regresji (równiej q^2 lub odpowiednio R^2) i części niewyjaśnionej (równiej $1-q^2$ lub odpowiednio $1-R^2$) za pomocą testu Fishera-Snedecora, porównując go dwie wariancje (patrz podrozdz. II.5.1.3).

Dodatek

Zasady korzystania z tablic statystycznych

W tablicach 1–10 podano kwantyle oraz wartości krytyczne najważniejszych rozkładów prawdopodobieństwa, których znajomość jest niezbędna do rozwiązywania zadań ze statystyki w zakresie, który obejmuje niniejszy podręcznik. Podano również wartości dystrybuanty standaryzowanego rozkładu normalnego. Oprócz standardowych zastosowań dystrybuanty, choćby do obliczania prawdopodobieństw w przedziałach określoności zmiennej losowej – tablica ta przydaje się również w sytuacjach, gdy tablica kwantyli tego rozkładu okazuje się zbyt mało dokładna. Korzystamy wówczas z własności:

$$\Phi(z(p)) = p.$$

Obecnie coraz częściej korzysta się z programów komputerowych do obliczania wartości dystrybuanty, kwantyli lub wartości krytycznych, jednakże podczas rozwiązywania zadań, wygodnie jest mieć „pod ręką” tablice w formie papierowej.

By ułatwić Czytelnikom korzystanie z tablic, przypominam, że dzięki symetrii względem pionowej osi układu współrzędnych, standaryzowanego rozkładu normalnego oraz rozkładu Studenta, wartości dystrybuanty i kwantyle tych rozkładów spełniają zależności:

$$\Phi(z) = 1 - \Phi(-z),$$

$$z(p) = -z(1-p),$$

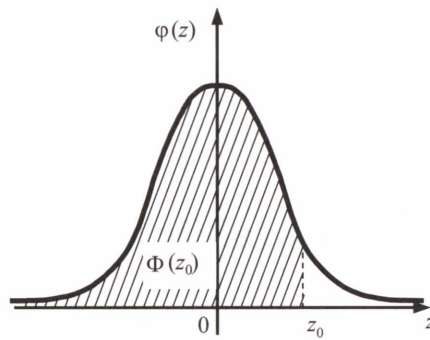
$$t(p, k) = -t(1-p, k).$$

Z wymienionych zależności korzystamy na przykład w sytuacji, gdy zachodzi konieczność obliczenia dystrybuanty dla ujemnych argumentów $z < 0$ lub gdy potrzebny jest kwantyl prawdopodobieństwa mniejszego od połowy $p < 0,5$.

Poniżej podano sposoby posługiwania się tablicami.

Tablica 1

Dwie pierwsze cyfry znaczące argumentu dystrybuanty znajdujemy w pierwszej kolumnie, a trzecią cyfrę znaczącą – w pierwszym wierszu. Na przecięciu wybranego wiersza i kolumny znajdujemy wartość dystrybuanty dla określonego argumentu (rys. D.1).

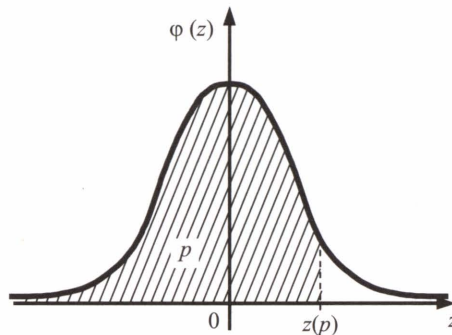


Rys. D.1. Interpretacja graficzna wartości dystrybuanty zmiennej losowej $Z \sim N(0,1)$

Zatem wartość dystrybuanty możemy przeczytać dla argumentu określonego z dokładnością do 0,01 (dwa miejsca po przecinku). Wartość ta odpowiada polu powierzchni pod wykresem funkcji gęstości, nad przedziałem $(-\infty, z_0)$. Na przykład: $\Phi(-1,05) = 1 - \Phi(1,05) = 1 - 0,85314 = 0,14686$ lub $\Phi(3,63) = 0,9^{3}858 = 0,999858$.

Tablica 2

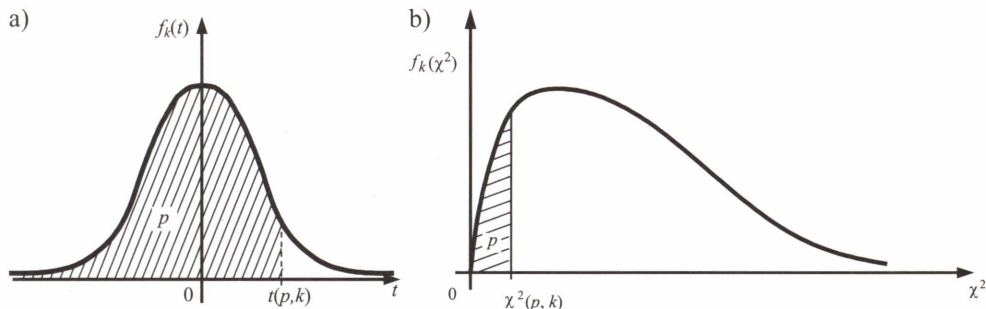
Zasada odczytywania wartości kwantylów (rys. D.2) jest analogiczna do zasad sformułowanych dla tablicy 1, z tym, że w tym wypadku dokładność argumentu kwantylu (prawdopodobieństwa p) wzrasta do 0,001. Na przykład: $z(0,637) = 0,3504$.



Rys. D.2. Interpretacja graficzna prawdopodobieństwa p dla kwantylu $z(p)$ w standaryzowanym rozkładzie normalnym $N(0,1)$

Tablice 3 i 4

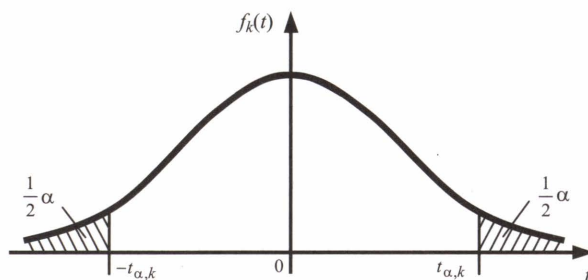
W pierwszym wierszu znajdujemy prawdopodobieństwo p , równe polu pod wykresem funkcji gęstości (rys. D.3), na lewo od szukanego kwantyla, a w pierwszej kolumnie – liczbę stopni swobody k , która jest parametrem charakteryzującym rozkład chi-kwadrat lub rozkład Studenta. Na przykład dla rozkładu $\chi^2(0,025; 7) = \chi^2_{0,975;7} = 1,690$ lub dla rozkładu Studenta: $t(0,999; 19) = 3,5794$.



Rys. D.3. Interpretacja graficzna prawdopodobieństwa p dla kwantylów $t(p, k)$ oraz $\chi^2(p, k)$ w rozkładach Studenta (a) oraz chi-kwadrat (b)

Tablica 5

W pierwszym wierszu znajdujemy prawdopodobieństwo α , równe podwojonemu polu pod wykresem funkcji gęstości (rys. D.4), na prawo od szukanej dodatniej wartości krytycznej, a w pierwszej kolumnie liczbę stopni swobody k , która jest parametrem charakteryzującym rozkład Studenta.



Rys. D.4. Interpretacja graficzna prawdopodobieństwa α dla wartości krytycznej $t_{\alpha, k}$ rozkładu Studenta T_k

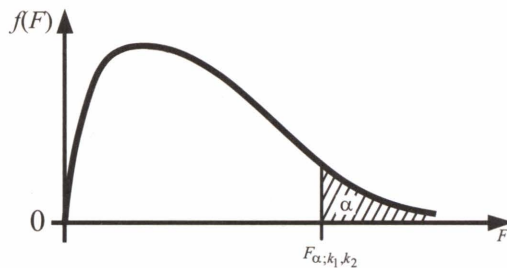
Warto pamiętać o relacji, jaka zachodzi pomiędzy kwantylami a wartościami krytycznymi rozkładu Studenta z tablic 4 i 5:

$$t_{\alpha, k} = t\left(1 - \frac{\alpha}{2}, k\right), \quad -t_{\alpha, k} = t\left(\frac{\alpha}{2}, k\right).$$

Na przykład: $t_{0,002;19} = t(0,999;19) = 3,5794$.

Tablice 6, 7 i 8

Każda z tych tablic jest w całości określona dla ustalonego poziomu istotności α (równego polu pod wykresem funkcji gęstości na prawo od szukanej wartości krytycznej) (rys. D.5), a pierwszy wiersz i pierwsza kolumna zawierają liczby stopni swobody, odpowiednio licznika i mianownika. Na przykład $F_{0,01;8,24} = 3,36$ czy $F_{0,05;100,21} = 1,88$.



Rys. D.5. Interpretacja graficzna prawdopodobieństwa α dla wartości krytycznej $F_{\alpha; k_1, k_2}$ rozkładu F-Sendocora

Mając wartości krytyczne $F_{\alpha; k_1, k_2}$ tego rozkładu, możemy w prosty sposób określić również jego kwantyle $F(p; k_1, k_2)$, na podstawie relacji:

$$F_{\alpha; k_1, k_2} = F(1 - \alpha; k_1, k_2).$$

Dodatkowo warto zapamiętać, że mając kwantyl dla zadanego prawdopodobieństwa p , możemy łatwo uzyskać kwantyl dla prawdopodobieństwa $1 - p$:

$$F(p; k_1, k_2) = \frac{1}{F(1 - p; k_1, k_2)}.$$

Tablica 9

Cała tablica podaje wartości graniczne przedziałów ufności dla wskaźnika struktury (frakcja) przy ustalonym poziomie ufności 0,95. W pierwszej kolumnie znajdujemy liczbę k elementów w n -elementowej próbie losowej, mających określoną cechę, natomiast w pierwszym wierszu – liczbę pozostałych elementów w próbie, czyli $n - k$. Na przecięciu wybranej kolumny i wiersza znajdują się wartości granic szukanego przedziału ufności.

Na przykład gdy w dwudziestoelementowej próbie losowej 9 elementów ma wyróżnioną cechę, to przedział ufności dla frakcji elementów o tej cesze, na poziomie ufności 0,95 jest następujący: (0,231; 0,685).

Tablica 10

Kwantyle $d_n(1 - \alpha)$ statystyki D_n Kołmogorowa (II.5.40) możemy odczytać dla trzech wybranych poziomów istotności α : 0,10; 0,05 i 0,01. Są to zatem kwantyle dla prawdopodobieństw $p = 1 - \alpha$, równych odpowiednio: 0,90; 0,95 i 0,99. Na przykład wykonując nieparametryczny test zgodności Kołmogorowa na podstawie próby losowej zawierającej 56 elementów, możemy otrzymać wartość kwantylu dla poziomu istotności $\alpha = 0,05$ wynoszącą $d_{56}(1 - 0,05) = d_{56}(0,95) = 0,178$.

TABLICE STATYSTYCZNE

Tablica 1

Wartości dystrybucyj standardyzowanego rozkładu normalnego $\Phi(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586	0,0
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56750	0,57142	0,57534	0,1
0,2	0,57926	0,58317	0,58706	0,59095	0,59484	0,59871	0,60257	0,60642	0,61026	0,61409	0,2
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173	0,3
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793	0,4
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240	0,5
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490	0,6
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524	0,7
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80510	0,80785	0,81057	0,81327	0,8
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891	0,9
1,0	0,84134	0,84375	0,84614	0,84850	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214	1,0
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298	1,1
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89616	0,89796	0,89973	0,90148	1,2
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774	1,3
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92786	0,92922	0,93056	0,93189	1,4
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408	1,5
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449	1,6
1,7	0,95544	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327	1,7
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062	1,8
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670	1,9
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169	2,0
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574	2,1
2,2	0,98610	0,98645	0,98679	0,98713	0,98746	0,98778	0,98809	0,98840	0,98870	0,98899	2,2
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158	2,3
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361	2,4

Tablica 1 cd.

2,5	0,99380	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520	2,5
2,6	0,99534	0,99547	0,99560	0,99573	0,99586	0,99598	0,99609	0,99621	0,99632	0,99643	2,6
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736	2,7
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807	2,8
2,9	0,99813	0,99819	0,99825	0,99830	0,99836	0,99841	0,99846	0,99851	0,99856	0,99860	2,9
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900	3,0
3,1	0,99903	0,99906	0,99910	0,99913	0,99915	0,99918	0,99921	0,99924	0,99926	0,99929	3,1
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950	3,2
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965	3,3
3,4	0,99966	0,99967	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976	3,4
3,5	0,9 ³ 767*	0,9 ³ 776	0,9 ³ 784	0,9 ³ 792	0,9 ³ 800	0,9 ³ 807	0,9 ³ 815	0,9 ³ 822	0,9 ³ 828	0,9 ³ 835	3,5
3,6	0,9 ³ 841	0,9 ³ 847	0,9 ³ 853	0,9 ³ 858	0,9 ³ 864	0,9 ³ 869	0,9 ³ 874	0,9 ³ 879	0,9 ³ 883	0,9 ³ 888	3,6
3,7	0,9 ³ 892	0,9 ³ 896	0,9 ⁴ 004	0,9 ⁴ 043	0,9 ⁴ 080	0,9 ⁴ 116	0,9 ⁴ 150	0,9 ⁴ 186	0,9 ⁴ 216	0,9 ⁴ 247	3,7
3,8	0,9 ⁴ 276	0,9 ⁴ 305	0,9 ⁴ 333	0,9 ⁴ 359	0,9 ⁴ 385	0,9 ⁴ 409	0,9 ⁴ 433	0,9 ⁴ 456	0,9 ⁴ 478	0,9 ⁴ 499	3,8
3,9	0,9 ⁴ 519	0,9 ⁴ 538	0,9 ⁴ 557	0,9 ⁴ 575	0,9 ⁴ 593	0,9 ⁴ 609	0,9 ⁴ 625	0,9 ⁴ 641	0,9 ⁴ 655	0,9 ⁴ 670	3,9
4,0	0,9 ⁴ 683	0,9 ⁴ 696	0,9 ⁴ 709	0,9 ⁴ 721	0,9 ⁴ 733	0,9 ⁴ 744	0,9 ⁴ 755	0,9 ⁴ 765	0,9 ⁴ 775	0,9 ⁴ 784	4,0
4,1	0,9 ⁴ 793	0,9 ⁴ 802	0,9 ⁴ 811	0,9 ⁴ 819	0,9 ⁴ 826	0,9 ⁴ 834	0,9 ⁴ 841	0,9 ⁴ 848	0,9 ⁴ 854	0,9 ⁴ 860	4,1
4,2	0,9 ⁴ 866	0,9 ⁴ 872	0,9 ⁴ 878	0,9 ⁴ 883	0,9 ⁴ 888	0,9 ⁴ 893	0,9 ⁴ 898	0,9 ⁵ 023	0,9 ⁵ 066	0,9 ⁵ 107	4,2
4,3	0,9 ⁵ 146	0,9 ⁵ 184	0,9 ⁵ 220	0,9 ⁵ 254	0,9 ⁵ 288	0,9 ⁵ 319	0,9 ⁵ 350	0,9 ⁵ 379	0,9 ⁵ 407	0,9 ⁵ 433	4,3
4,4	0,9 ⁵ 459	0,9 ⁵ 483	0,9 ⁵ 506	0,9 ⁵ 529	0,9 ⁵ 550	0,9 ⁵ 571	0,9 ⁵ 590	0,9 ⁵ 609	0,9 ⁵ 627	0,9 ⁵ 644	4,4
4,5	0,9 ⁵ 660	0,9 ⁵ 676	0,9 ⁵ 691	0,9 ⁵ 705	0,9 ⁵ 719	0,9 ⁵ 732	0,9 ⁵ 744	0,9 ⁵ 756	0,9 ⁵ 768	0,9 ⁵ 778	4,5
4,6	0,9 ⁵ 789	0,9 ⁵ 799	0,9 ⁵ 808	0,9 ⁵ 817	0,9 ⁵ 826	0,9 ⁵ 834	0,9 ⁵ 842	0,9 ⁵ 849	0,9 ⁵ 857	0,9 ⁵ 863	4,6
4,7	0,9 ⁵ 870	0,9 ⁵ 876	0,9 ⁵ 882	0,9 ⁵ 888	0,9 ⁵ 893	0,9 ⁵ 898	0,9 ⁶ 032	0,9 ⁶ 079	0,9 ⁶ 124	0,9 ⁶ 166	4,7
4,8	0,9 ⁶ 207	0,9 ⁶ 245	0,9 ⁶ 282	0,9 ⁶ 317	0,9 ⁶ 351	0,9 ⁶ 383	0,9 ⁶ 413	0,9 ⁶ 442	0,9 ⁶ 470	0,9 ⁶ 496	4,8
4,9	0,9 ⁶ 521	0,9 ⁶ 545	0,9 ⁶ 567	0,9 ⁶ 589	0,9 ⁶ 609	0,9 ⁶ 629	0,9 ⁶ 648	0,9 ⁶ 665	0,9 ⁶ 682	0,9 ⁶ 698	4,9

* Zapis skrócony, np. 0,9³ 797 = 0,999797.

Tablica 2
Kwantyle standaryzowanego rozkładu normalnego $z(p)$

p	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	p
0,50	0,0000	0,0025	0,0050	0,0075	0,0100	0,0125	0,0150	0,0175	0,0200	0,0226	0,50
0,51	0,0251	0,0276	0,0308	0,0326	0,0351	0,0376	0,4012	0,0426	0,0451	0,0476	0,51
0,52	0,0501	0,0527	0,0552	0,0577	0,0602	0,0627	0,0652	0,0677	0,0702	0,0728	0,52
0,53	0,0753	0,0778	0,0803	0,0828	0,0853	0,0878	0,0904	0,0929	0,0954	0,0979	0,53
0,54	0,1004	0,1029	0,1055	0,1080	0,1105	0,1130	0,1156	0,1181	0,1206	0,1231	0,54
0,55	0,1257	0,1282	0,1307	0,1332	0,1358	0,1383	0,1408	0,1434	0,1459	0,1484	0,55
0,56	0,1510	0,1535	0,1560	0,1586	0,1611	0,1637	0,1662	0,1687	0,1713	0,1738	0,56
0,57	0,1764	0,1789	0,1815	0,1840	0,1866	0,1891	0,1917	0,1942	0,1968	0,1993	0,57
0,58	0,2019	0,2044	0,2070	0,2096	0,2121	0,2147	0,2173	0,2198	0,2224	0,2250	0,58
0,59	0,2275	0,2301	0,2327	0,2353	0,2378	0,2404	0,2430	0,2456	0,2482	0,2508	0,59
0,60	0,2533	0,2559	0,2585	0,2611	0,2637	0,2663	0,2689	0,2715	0,2741	0,2767	0,60
0,61	0,2793	0,2819	0,2845	0,2871	0,2898	0,2924	0,2950	0,2976	0,3002	0,3028	0,61
0,62	0,3055	0,3081	0,3107	0,3134	0,3160	0,3186	0,3213	0,3239	0,3266	0,3292	0,62
0,63	0,3318	0,3345	0,3371	0,3398	0,3425	0,3451	0,3478	0,3504	0,3531	0,3558	0,63
0,64	0,3585	0,3611	0,3638	0,3665	0,3692	0,3719	0,3745	0,3772	0,3799	0,3826	0,64
0,65	0,3853	0,3880	0,3907	0,3934	0,3961	0,3988	0,4016	0,4043	0,4070	0,4097	0,65
0,66	0,4125	0,4152	0,4179	0,4207	0,4234	0,4261	0,4289	0,4316	0,4344	0,4371	0,66
0,67	0,4399	0,4427	0,4454	0,4482	0,4510	0,4538	0,4565	0,4593	0,4621	0,4649	0,67
0,68	0,4677	0,4705	0,4733	0,4761	0,4789	0,4817	0,4845	0,4874	0,4902	0,4930	0,68
0,69	0,4958	0,4987	0,5015	0,5044	0,5072	0,5101	0,5129	0,5158	0,5187	0,5215	0,69
0,70	0,5244	0,5273	0,5302	0,5330	0,5359	0,5388	0,5417	0,5446	0,5475	0,5505	0,70
0,71	0,5534	0,5563	0,5592	0,5622	0,5651	0,5680	0,5710	0,5739	0,5769	0,5799	0,71
0,72	0,5828	0,5858	0,5888	0,5918	0,5948	0,5978	0,6008	0,6038	0,6068	0,6098	0,72
0,73	0,6128	0,6158	0,6189	0,6219	0,6250	0,6280	0,6311	0,6341	0,6372	0,6403	0,73
0,74	0,6433	0,6464	0,6495	0,6526	0,6557	0,6588	0,6619	0,6651	0,6682	0,6713	0,74

Tablica 2 cd.

0,75	0,6745	0,6776	0,6808	0,6840	0,6871	0,6903	0,6935	0,6967	0,6999	0,7031	0,75
0,76	0,7063	0,7095	0,7127	0,7160	0,7192	0,7225	0,7257	0,7290	0,7323	0,7356	0,76
0,77	0,7388	0,7421	0,7454	0,7488	0,7521	0,7554	0,7587	0,7621	0,7655	0,7688	0,77
0,78	0,7722	0,7756	0,7790	0,7824	0,7858	0,7892	0,7926	0,7960	0,7995	0,8030	0,78
0,79	0,8064	0,8099	0,8134	0,8169	0,8204	0,8239	0,8274	0,8309	0,8345	0,8380	0,79
0,80	0,8416	0,8452	0,8488	0,8524	0,8560	0,8596	0,8632	0,8669	0,8705	0,8742	0,80
0,81	0,8779	0,8816	0,8853	0,8890	0,8927	0,8965	0,9002	0,9040	0,9078	0,9116	0,81
0,82	0,9154	0,9192	0,9230	0,9269	0,9307	0,9346	0,9385	0,9424	0,9463	0,9502	0,82
0,83	0,9542	0,9581	0,9621	0,9661	0,9701	0,9741	0,9781	0,9822	0,9863	0,9904	0,83
0,84	0,9945	0,9986	1,0027	1,0070	1,0110	1,0152	1,0194	1,0236	1,0279	1,0321	0,84
0,85	1,0364	1,0407	1,0450	1,0494	1,0537	1,0581	1,0625	1,0669	1,0714	1,0758	0,85
0,86	1,0803	1,0848	1,0893	1,0939	1,0985	1,1031	1,1077	1,1123	1,1170	1,1217	0,86
0,87	1,1264	1,1311	1,1359	1,1407	1,1455	1,1503	1,1552	1,1601	1,1650	1,1700	0,87
0,88	1,1750	1,1800	1,1850	1,1901	1,1952	1,2004	1,2055	1,2107	1,2160	1,2212	0,88
0,89	1,2265	1,2319	1,2372	1,2426	1,2481	1,2536	1,2591	1,2646	1,2702	1,2759	0,89
0,90	1,2815	1,2873	1,2930	1,2988	1,3047	1,3106	1,3165	1,3225	1,3285	1,3346	0,90
0,91	1,3407	1,3469	1,3532	1,3595	1,3658	1,3722	1,3787	1,3852	1,3917	1,3984	0,91
0,92	1,4051	1,4118	1,4186	1,4255	1,4325	1,4395	1,4466	1,4538	1,4611	1,4684	0,92
0,93	1,4758	1,4833	1,4908	1,4985	1,5063	1,5141	1,5220	1,5301	1,5382	1,5464	0,93
0,94	1,5548	1,5632	1,5718	1,5805	1,5893	1,5982	1,6072	1,6164	1,6258	1,6352	0,94
0,95	1,6448	1,6546	1,6646	1,6747	1,6849	1,6954	1,7060	1,7169	1,7279	1,7392	0,95
0,96	1,7507	1,7624	1,7744	1,7866	1,7991	1,8119	1,8250	1,8384	1,8522	1,8663	0,96
0,97	1,8808	1,8957	1,9110	1,9268	1,9431	1,9600	1,9774	1,9954	2,0141	2,0335	0,97
0,98	2,0537	2,0748	2,0969	2,1201	2,1444	2,1701	2,1973	2,2262	2,2571	2,2904	0,98
0,99	2,3263	2,3656	2,4089	2,4573	2,5121	2,5758	2,6521	2,7478	2,8782	3,0902	0,99

Tablica 3 – część I
Kwantyle rozkładu chi-kwadrat $\chi^2(p, k)$ (równe wartościom krytycznym $\chi^2_{\alpha=1-p, k}$)

$k \backslash p$	0,0005	0,001	0,005	0,01	0,025	0,03	0,05	0,10	0,20	0,30	0,40	$p \backslash k$
1	0,0 ⁶ 393	0,0 ⁵ 157	0,0 ⁴ 393	0,0 ³ 157	0,0 ³ 982	0,001	0,0 ³ 393	0,0158	0,0642	0,148	0,275	1
2	0,0 ² 100	0,0 ² 200	0,0100	0,0201	0,0506	0,061	0,103	0,211	0,446	0,713	1,022	2
3	0,0153	0,0243	0,0717	0,115	0,216	0,245	0,352	0,584	1,005	1,424	1,869	3
4	0,0639	0,0908	0,207	0,297	0,484	0,535	0,711	1,064	1,649	2,195	2,753	4
5	0,158	0,210	0,412	0,554	0,831	0,903	1,145	1,610	2,343	3,000	3,655	5
6	0,299	0,381	0,676	0,872	1,237	1,330	1,635	2,204	3,070	3,828	4,570	6
7	0,485	0,598	0,989	1,239	1,690	1,802	2,167	2,833	3,822	4,671	5,493	7
8	0,710	0,857	1,344	1,646	2,180	2,310	2,733	3,490	4,594	5,527	6,423	8
9	0,972	1,153	1,735	2,088	2,700	2,848	3,325	4,168	5,380	6,393	7,357	9
10	1,265	1,479	2,156	2,558	3,247	3,412	3,940	4,865	6,179	7,267	8,295	10
11	1,587	1,834	2,603	3,053	3,816	3,997	4,575	5,578	6,989	8,148	9,237	11
12	1,934	2,214	3,074	3,571	4,404	4,601	5,226	6,304	7,807	9,034	10,182	12
13	2,305	2,617	3,565	4,107	5,009	5,221	5,892	7,042	8,634	9,926	11,129	13
14	2,697	3,041	4,075	4,660	5,629	5,856	6,571	7,790	9,467	10,821	12,079	14
15	3,108	3,483	4,601	5,229	6,262	6,503	7,261	8,547	10,307	11,721	13,030	15
16	3,536	3,942	5,142	5,812	6,908	7,163	7,962	9,312	11,152	12,624	13,983	16
17	3,980	4,416	5,697	6,408	7,564	7,832	8,672	10,085	12,002	13,531	14,937	17
18	4,439	4,905	6,265	7,015	8,231	8,512	9,390	10,865	12,857	14,440	15,893	18
19	4,912	5,407	6,844	7,633	8,907	9,200	10,117	11,651	13,716	15,352	16,850	19
20	5,398	5,921	7,343	8,260	9,591	9,897	10,851	12,443	14,578	16,266	17,809	20
21	5,896	6,447	8,034	8,897	10,283	10,601	11,591	13,240	15,445	17,182	18,768	21
22	6,404	6,983	8,643	9,542	10,982	11,313	12,338	14,041	16,314	18,101	19,729	22
23	6,924	7,529	9,260	10,196	11,688	12,030	13,091	14,848	17,187	19,021	20,690	23
24	7,453	8,085	9,886	10,856	12,401	12,754	13,848	15,659	18,062	19,943	21,652	24
25	7,991	8,649	10,520	11,524	13,120	13,484	14,611	16,473	18,940	20,867	22,616	25

Tablica 3 – część I cd.

26	8,538	9,222	11,160	12,198	13,844	14,219	15,379	17,292	19,820	21,792	23,579	26
27	9,093	9,803	11,808	12,879	14,573	14,959	16,151	18,114	20,703	22,719	24,544	27
28	9,656	10,391	12,461	13,565	15,308	15,704	16,928	18,939	21,588	23,647	25,509	28
29	10,227	10,986	13,121	14,256	16,047	16,454	17,708	19,768	22,475	24,577	26,475	29
30	10,804	11,588	13,787	14,953	16,791	17,208	18,493	20,599	23,364	25,508	27,442	30
31	11,389	12,196	14,458	15,655	17,539	17,966	19,281	21,434	24,255	26,440	28,409	31
32	11,979	12,811	15,134	16,362	18,291	18,727	20,072	22,271	25,148	27,373	29,376	32
33	12,576	13,431	15,815	17,073	19,047	19,493	20,867	23,110	26,042	28,307	30,344	33
34	13,176	14,057	16,501	17,789	19,806	20,262	21,664	23,952	26,938	29,242	31,313	34
35	13,788	14,688	17,192	18,509	20,569	21,035	22,465	24,797	27,836	30,178	32,282	35
36	14,401	15,324	17,887	19,233	21,336	21,811	23,269	25,643	28,735	31,115	33,252	36
37	15,020	15,965	18,586	19,960	22,106	22,589	24,075	26,492	29,635	32,053	34,222	37
38	15,644	16,611	19,289	20,691	22,878	23,371	24,884	27,343	30,537	32,992	35,192	38
39	16,273	17,262	19,996	21,426	23,654	24,156	25,695	28,196	31,441	33,932	36,163	39
40	16,906	17,916	20,707	22,164	24,433	24,944	26,509	29,051	32,345	34,872	37,134	40
41	17,544	18,575	21,421	22,906	25,215	25,734	27,326	29,907	33,251	35,813	38,105	41
42	18,186	19,238	22,138	23,650	25,909	26,527	28,144	30,765	34,157	36,755	39,077	42
43	18,832	19,905	22,859	24,398	26,785	27,322	28,965	31,625	35,065	37,698	40,050	43
44	19,482	20,576	23,584	25,148	27,575	28,119	29,787	32,487	35,974	38,641	41,022	44
45	20,137	21,251	24,311	25,901	28,366	28,919	30,612	33,350	36,884	39,585	41,955	45
46	20,794	21,929	25,041	26,657	29,160	29,722	31,439	34,215	37,795	40,529	42,968	46
47	21,456	22,610	25,775	27,416	29,956	30,526	32,268	35,081	38,708	41,474	43,942	47
48	22,121	23,295	26,511	28,177	30,755	31,332	33,098	35,949	39,621	42,420	44,915	48
49	22,789	23,983	27,249	28,941	31,555	32,141	33,930	36,818	40,534	43,366	45,889	49
50	23,461	24,674	27,991	29,707	32,357	32,951	34,764	37,689	41,449	44,313	46,864	50
51	24,136	25,368	28,735	30,475	33,162	33,763	35,600	38,560	42,365	45,261	47,838	51
52	24,814	26,065	29,481	31,246	33,968	34,577	36,437	39,433	43,281	46,209	48,813	52
53	25,495	26,765	30,230	32,018	34,776	35,393	37,276	40,308	44,199	47,157	49,788	53
54	26,179	27,468	30,981	32,793	35,586	36,211	38,116	41,183	45,117	48,106	50,764	54
55	26,866	28,173	31,735	33,570	36,398	37,030	38,958	42,060	46,036	49,054	51,739	55

Tablica 3 – część II
Kwantyle rozkładu chi-kwadrat $\chi^2(p, k)$ (równe wartościom krytycznym $\chi^2_{\alpha=1-p, k}$)

$k \backslash p$	0,50	0,60	0,70	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	$p \backslash k$
1	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,635	7,879	10,828	12,116	1
2	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202	2
3	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730	3
4	3,357	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,997	4
5	4,351	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105	5
6	5,348	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103	6
7	6,346	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018	7
8	7,344	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,124	27,868	8
9	8,343	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666	9
10	9,342	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,420	10
11	10,341	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136	11
12	11,340	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821	12
13	12,340	13,636	15,119	16,985	19,812	22,362	24,735	27,688	29,819	34,528	36,478	13
14	13,339	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109	14
15	14,339	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719	15
16	15,338	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308	16
17	16,338	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879	17
18	17,338	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434	18
19	18,338	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973	19
20	19,337	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498	20
21	20,337	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797	49,010	21
22	21,337	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268	50,511	22
23	22,337	24,069	26,018	28,429	32,007	35,172	38,086	41,638	44,181	49,728	52,000	23
24	23,337	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179	53,479	24
25	24,337	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,620	54,947	25

Tablica 3 – część II cd.

26	25,336	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052	56,407	26
27	26,336	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476	57,858	27
28	27,336	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892	59,300	28
29	28,336	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,301	60,735	29
30	29,336	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703	62,162	30
31	30,336	32,349	34,598	37,359	41,422	44,985	48,232	52,191	55,003	61,098	63,582	31
32	31,336	33,381	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487	64,995	32
33	32,336	34,413	36,731	39,572	43,745	47,400	50,725	54,776	57,648	63,870	66,403	33
34	33,336	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247	67,803	34
35	34,336	36,475	38,859	41,778	46,059	49,802	53,203	57,342	60,275	66,619	69,199	35
36	35,336	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985	70,588	36
37	36,336	38,535	40,984	43,978	48,363	52,192	55,668	59,893	62,883	69,346	71,972	37
38	37,335	39,564	42,045	45,076	49,513	53,384	56,896	61,162	64,181	70,703	73,351	38
39	38,335	40,593	43,105	46,173	50,660	54,572	58,120	62,428	65,476	72,055	74,725	39
40	39,335	41,622	44,165	47,269	51,805	55,758	59,342	63,691	66,766	73,402	76,095	40
41	40,335	42,651	45,224	48,363	52,949	56,942	60,561	64,950	68,053	74,745	77,459	41
42	41,335	43,679	46,282	49,456	54,090	58,124	61,777	66,206	69,336	76,084	78,820	42
43	42,335	44,706	47,339	50,548	55,230	59,304	62,990	67,459	70,616	77,419	80,176	43
44	43,335	45,734	48,396	51,639	56,369	60,481	64,201	68,710	71,893	78,750	81,528	44
45	44,335	46,761	49,452	52,729	57,505	61,656	65,410	69,957	73,166	80,077	82,876	45
46	45,335	47,787	50,507	53,818	58,641	62,830	66,617	71,201	74,437	81,400	84,220	46
47	46,335	48,814	51,562	54,906	59,774	64,001	67,821	72,443	75,704	82,720	85,560	47
48	47,335	49,840	52,616	55,993	60,907	65,171	69,023	73,683	76,969	84,037	86,897	48
49	48,335	50,866	53,670	57,079	62,038	66,339	70,222	74,919	78,231	85,351	88,231	49
50	49,335	51,892	54,723	58,164	63,167	67,505	71,420	76,154	79,490	86,661	89,561	50
51	50,335	52,917	55,775	59,248	64,295	68,669	72,616	77,386	80,747	87,968	90,887	51
52	51,335	53,942	56,827	60,332	65,422	69,832	73,810	78,616	82,001	89,272	92,211	52
53	52,335	54,967	57,879	61,414	66,548	70,993	75,002	79,843	83,253	90,573	93,531	53
54	53,335	55,992	58,930	62,496	67,673	72,153	76,192	81,069	84,502	91,872	94,849	54
55	54,335	57,016	59,980	63,577	68,796	73,311	77,380	82,292	85,749	93,168	96,163	55

Tablica 4
Kwantyle rozkładu T-Studenta $t(p, k)$

$k \backslash p$	0,60	0,70	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	$p \backslash k$
1	0,3249	0,7265	1,3764	3,0777	6,3138	12,7062	31,8205	63,6567	318,309	636,612	1
2	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9646	9,9248	22,3271	31,5991	2
3	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8409	10,2145	12,9240	3
4	0,2707	0,5686	0,9410	1,5332	2,1318	2,7764	3,7469	4,6041	7,1732	8,6103	4
5	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8934	6,8688	5
6	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2076	5,9588	6
7	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9980	3,4995	4,7853	5,4079	7
8	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008	5,0413	8
9	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809	9
10	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5869	10
11	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370	11
12	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178	12
13	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208	13
14	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405	14
15	0,2579	0,5357	0,8662	1,3406	1,7531	2,1314	2,6025	2,9467	3,7328	4,0728	15
16	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150	16
17	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651	17
18	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216	18
19	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834	19
20	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495	20
21	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193	21
22	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921	22
23	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676	23
24	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454	24
25	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251	25

Tablica 4 cd.

26	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066	26
27	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896	27
28	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739	28
29	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594	29
30	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460	30
31	0,2555	0,5298	0,8534	1,3095	1,6955	2,0395	2,4528	2,7440	3,3749	3,6335	31
32	0,2555	0,5297	0,8530	1,3086	1,6939	2,0369	2,4487	2,7385	3,3653	3,6218	32
33	0,2554	0,5295	0,8526	1,3077	1,6924	2,0345	2,4448	2,7333	3,3563	3,6109	33
34	0,2553	0,5294	0,8523	1,3070	1,6909	2,0322	2,4411	2,7284	3,3479	3,6007	34
35	0,2553	0,5292	0,8520	1,3062	1,6896	2,0301	2,4377	2,7238	3,3400	3,5911	35
36	0,2552	0,5291	0,8517	1,3055	1,6883	2,0281	2,4345	2,7195	3,3326	3,5821	36
37	0,2552	0,5289	0,8514	1,3049	1,6871	2,0262	2,4314	2,7154	3,3256	3,5737	37
38	0,2551	0,5288	0,8512	1,3042	1,6860	2,0244	2,4286	2,7116	3,3190	3,5657	38
39	0,2551	0,5287	0,8509	1,3036	1,6849	2,0227	2,4258	2,7079	3,3128	3,5581	39
40	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510	40
41	0,2550	0,5285	0,8505	1,3025	1,6829	2,0195	2,4208	2,7012	3,3013	3,5442	41
42	0,2550	0,5284	0,8503	1,3020	1,6820	2,0181	2,4185	2,6981	3,2960	3,5377	42
43	0,2549	0,5283	0,8501	1,3016	1,6811	2,0167	2,4163	2,6951	3,2909	3,5316	43
44	0,2549	0,5282	0,8499	1,3011	1,6802	2,0154	2,4141	2,6923	3,2861	3,5258	44
45	0,2549	0,5281	0,8497	1,3006	1,6794	2,0141	2,4121	2,6896	3,2815	3,5203	45
46	0,2548	0,5281	0,8495	1,3002	1,6787	2,0129	2,4102	2,6870	3,2771	3,5150	46
47	0,2548	0,5280	0,8493	1,2998	1,6779	2,0117	2,4083	2,6846	3,2729	3,5099	47
48	0,2548	0,5279	0,8492	1,2994	1,6772	2,0106	2,4066	2,6822	3,2689	3,5051	48
49	0,2547	0,5278	0,8490	1,2991	1,6766	2,0096	2,4049	2,6800	3,2651	3,5004	49
50	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960	50

Tablica 5
Wartości krytyczne rozkładu T-Studenta $t_{\alpha,k}$

α k	0,50	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001	α k
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192	1
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991	2
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240	3
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103	4
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688	5
6	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588	6
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	4,7853	5,4079	7
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413	8
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2968	4,7809	9
10	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869	10
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370	11
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178	12
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208	13
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7874	4,1405	14
15	0,6912	1,3406	1,7531	2,1314	2,6025	2,9467	3,2860	3,7328	4,0728	15
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150	16
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651	17
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216	18
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834	19
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495	20
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193	21
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921	22
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676	23
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7454	24
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251	25

Tablica 5 cd.

26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066	26
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896	27
28	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739	28
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594	29
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460	30
32	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,3653	3,6218	32
34	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,3479	3,6007	34
36	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,3326	3,5821	36
38	0,6810	1,3042	1,6860	2,0244	2,4286	2,7116	2,9803	3,3190	3,5657	38
40	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510	40
42	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,2960	3,5377	42
44	0,6801	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,2861	3,5258	44
46	0,6799	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,2771	3,5150	46
48	0,6796	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,2689	3,5051	48
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,2614	3,4960	50
55	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682	2,9247	3,2451	3,4764	55
60	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602	60
65	0,6783	1,2947	1,6686	1,9971	2,3851	2,6536	2,9060	3,2204	3,4466	65
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,2108	3,4350	70
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163	80
90	0,6772	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,1833	3,4019	90
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,1737	3,3905	100
120	0,6765	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735	120
150	0,6761	1,2872	1,6551	1,9759	2,3515	2,6090	2,8492	3,1455	3,3566	150
200	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006	2,8385	3,1315	3,3398	200
250	0,6755	1,2849	1,6510	1,9695	2,3414	2,5956	2,8322	3,1232	3,3299	250

Tablica 6 – część I
Wartości krytyczne rozkładu F-Snedecora dla poziomu istotności 0,01 $F_{0,01;k_1,k_2}$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	$k_1 \backslash k_2$
1	4052,2	4999,5	5403,4	5624,6	5763,7	5859,0	5928,4	5981,1	6022,5	6055,9	6106,3	6142,7	6170,1	6191,5	6208,7	1
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,44	99,44	99,45	2
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,92	26,83	26,75	26,69	3
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,25	14,15	14,08	14,02	4
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,77	9,68	9,61	9,55	5
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,60	7,52	7,45	7,40	6
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,36	6,28	6,21	6,16	7
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,56	5,48	5,41	5,36	8
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	5,01	4,92	4,86	4,81	9
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,60	4,52	4,46	4,41	10
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,29	4,21	4,15	4,10	11
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,05	3,97	3,91	3,86	12
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,86	3,78	3,72	3,66	13
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,70	3,62	3,56	3,51	14
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,56	3,49	3,42	3,37	15
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,45	3,37	3,31	3,26	16
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,35	3,27	3,21	3,16	17
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,27	3,19	3,13	3,08	18
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,19	3,12	3,05	3,00	19
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,13	3,05	2,99	2,94	20
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,07	2,99	2,93	2,88	21
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	3,02	2,94	2,88	2,83	22
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,97	2,89	2,83	2,78	23
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,93	2,85	2,79	2,74	24
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,89	2,81	2,75	2,70	25
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,86	2,78	2,72	2,66	26
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,82	2,75	2,68	2,63	27
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,79	2,72	2,65	2,60	28
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,77	2,69	2,63	2,57	29
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,74	2,66	2,60	2,55	30

Tablica 6 – część I cd.

32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,80	2,70	2,62	2,55	2,50	32
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,76	2,66	2,58	2,51	2,46	34
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,72	2,62	2,54	2,48	2,43	36
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,69	2,59	2,51	2,45	2,40	38
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,56	2,48	2,42	2,37	40
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,64	2,54	2,46	2,40	2,34	42
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,62	2,52	2,44	2,37	2,32	44
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,60	2,50	2,42	2,35	2,30	46
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,58	2,48	2,40	2,33	2,28	48
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,56	2,46	2,38	2,32	2,27	50
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,53	2,42	2,34	2,28	2,23	55
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,39	2,31	2,25	2,20	60
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	2,47	2,37	2,29	2,23	2,17	65
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,45	2,35	2,27	2,20	2,15	70
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,42	2,31	2,23	2,17	2,12	80
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,39	2,29	2,21	2,14	2,09	90
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,37	2,27	2,19	2,12	2,07	100
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,23	2,15	2,09	2,03	120
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,31	2,20	2,12	2,06	2,00	150
180	6,78	4,73	3,89	3,43	3,12	2,90	2,74	2,61	2,51	2,42	2,28	2,18	2,10	2,04	1,98	180
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,27	2,17	2,09	2,03	1,97	200
250	6,74	4,69	3,86	3,40	3,09	2,87	2,71	2,58	2,48	2,39	2,26	2,15	2,07	2,01	1,95	250
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,24	2,14	2,06	1,99	1,94	300
350	6,71	4,67	3,84	3,37	3,07	2,85	2,69	2,56	2,46	2,37	2,24	2,13	2,05	1,99	1,93	350
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	2,45	2,37	2,23	2,13	2,05	1,98	1,92	400
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,22	2,12	2,04	1,97	1,92	500
600	6,68	4,64	3,81	3,35	3,05	2,83	2,67	2,54	2,44	2,35	2,21	2,11	2,03	1,96	1,91	600
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35	2,21	2,11	2,03	1,96	1,90	700
800	6,67	4,63	3,81	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,21	2,10	2,02	1,96	1,90	800
999	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,20	2,10	2,02	1,95	1,90	999

Tablica 6 – część II
Wartości krytyczne rozkładu F-Snedecora dla poziomu istotności 0,01 $F_{0,01;k_1,k_2}$

$k_2 \backslash k_1$	22	24	26	28	30	35	40	45	50	60	80	100	200	500	999	$k_1 \backslash k_2$
1	6222,8	6234,6	6244,6	6253,2	6260,7	6275,6	6286,8	6295,5	6302,5	6313,0	6326,2	6334,1	6350,0	6359,5	6362,7	1
2	99,45	99,46	99,46	99,46	99,47	99,47	99,47	99,48	99,48	99,48	99,49	99,49	99,49	99,50	99,50	2
3	26,64	26,60	26,56	26,53	26,50	26,45	26,41	26,38	26,35	26,32	26,27	26,24	26,18	26,15	26,14	3
4	13,97	13,93	13,89	13,86	13,84	13,79	13,75	13,71	13,69	13,65	13,61	13,58	13,52	13,49	13,47	4
5	9,51	9,47	9,43	9,40	9,38	9,33	9,29	9,26	9,24	9,20	9,16	9,13	9,08	9,04	9,03	5
6	7,35	7,31	7,28	7,25	7,23	7,18	7,14	7,11	7,09	7,06	7,01	6,99	6,93	6,90	6,89	6
7	6,11	6,07	6,04	6,02	5,99	5,94	5,91	5,88	5,86	5,82	5,78	5,75	5,70	5,67	5,66	7
8	5,32	5,28	5,25	5,22	5,20	5,15	5,12	5,09	5,07	5,03	4,99	4,96	4,91	4,88	4,87	8
9	4,77	4,73	4,70	4,67	4,65	4,60	4,57	4,54	4,52	4,48	4,44	4,41	4,36	4,33	4,32	9
10	4,36	4,33	4,30	4,27	4,25	4,20	4,17	4,14	4,12	4,08	4,04	4,01	3,96	3,93	3,92	10
11	4,06	4,02	3,99	3,96	3,94	3,89	3,86	3,83	3,81	3,78	3,73	3,71	3,66	3,62	3,61	11
12	3,82	3,78	3,75	3,72	3,70	3,65	3,62	3,59	3,57	3,54	3,49	3,47	3,41	3,38	3,37	12
13	3,62	3,59	3,56	3,53	3,51	3,46	3,43	3,40	3,38	3,34	3,30	3,27	3,22	3,19	3,18	13
14	3,46	3,43	3,40	3,37	3,35	3,30	3,27	3,24	3,22	3,18	3,14	3,11	3,06	3,03	3,02	14
15	3,33	3,29	3,26	3,24	3,21	3,17	3,13	3,10	3,08	3,05	3,00	2,98	2,92	2,89	2,88	15
16	3,22	3,18	3,15	3,12	3,10	3,05	3,02	2,99	2,97	2,93	2,89	2,86	2,81	2,78	2,76	16
17	3,12	3,08	3,05	3,03	3,00	2,96	2,92	2,89	2,87	2,83	2,79	2,76	2,71	2,68	2,66	17
18	3,03	3,00	2,97	2,94	2,92	2,87	2,84	2,81	2,78	2,75	2,70	2,68	2,62	2,59	2,58	18
19	2,96	2,92	2,89	2,87	2,84	2,80	2,76	2,73	2,71	2,67	2,63	2,60	2,55	2,51	2,50	19
20	2,90	2,86	2,83	2,80	2,78	2,73	2,69	2,67	2,64	2,61	2,56	2,54	2,48	2,44	2,43	20
21	2,84	2,80	2,77	2,74	2,72	2,67	2,64	2,61	2,58	2,55	2,50	2,48	2,42	2,38	2,37	21
22	2,78	2,75	2,72	2,69	2,67	2,62	2,58	2,55	2,53	2,50	2,45	2,42	2,36	2,33	2,32	22
23	2,74	2,70	2,67	2,64	2,62	2,57	2,54	2,51	2,48	2,45	2,40	2,37	2,32	2,28	2,27	23
24	2,70	2,66	2,63	2,60	2,58	2,53	2,49	2,46	2,44	2,40	2,36	2,33	2,27	2,24	2,22	24
25	2,66	2,62	2,59	2,56	2,54	2,49	2,45	2,42	2,40	2,36	2,32	2,29	2,23	2,19	2,18	25
26	2,62	2,58	2,55	2,53	2,50	2,45	2,42	2,39	2,36	2,33	2,28	2,25	2,19	2,16	2,14	26
27	2,59	2,55	2,52	2,49	2,47	2,42	2,38	2,35	2,33	2,29	2,25	2,22	2,16	2,12	2,11	27
28	2,56	2,52	2,49	2,46	2,44	2,39	2,35	2,32	2,30	2,26	2,22	2,19	2,13	2,09	2,08	28
29	2,53	2,49	2,46	2,44	2,41	2,36	2,33	2,30	2,27	2,23	2,19	2,16	2,10	2,06	2,05	29
30	2,51	2,47	2,44	2,41	2,39	2,34	2,30	2,27	2,25	2,21	2,16	2,13	2,07	2,03	2,02	30

Tablica 6 – część II cd.

32	2,46	2,42	2,39	2,36	2,34	2,29	2,25	2,22	2,20	2,16	2,11	2,08	2,02	1,98	1,97	32
34	2,42	2,38	2,35	2,32	2,30	2,25	2,21	2,18	2,16	2,12	2,07	2,04	1,98	1,94	1,92	34
36	2,38	2,35	2,32	2,29	2,26	2,21	2,18	2,14	2,12	2,08	2,03	2,00	1,94	1,90	1,89	36
38	2,35	2,32	2,28	2,26	2,23	2,18	2,14	2,11	2,09	2,05	2,00	1,97	1,90	1,86	1,85	38
40	2,33	2,29	2,26	2,23	2,20	2,15	2,11	2,08	2,06	2,02	1,97	1,94	1,87	1,83	1,82	40
42	2,30	2,26	2,23	2,20	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,91	1,85	1,80	1,79	42
44	2,28	2,24	2,21	2,18	2,15	2,10	2,07	2,03	2,01	1,97	1,92	1,89	1,82	1,78	1,76	44
46	2,26	2,22	2,19	2,16	2,13	2,08	2,04	2,01	1,99	1,95	1,90	1,86	1,80	1,76	1,74	46
48	2,24	2,20	2,17	2,14	2,12	2,06	2,02	1,99	1,97	1,93	1,88	1,84	1,78	1,73	1,72	48
50	2,22	2,18	2,15	2,12	2,10	2,05	2,01	1,97	1,95	1,91	1,86	1,82	1,76	1,71	1,70	50
55	2,18	2,15	2,11	2,08	2,06	2,01	1,97	1,94	1,91	1,87	1,82	1,78	1,71	1,67	1,65	55
60	2,15	2,12	2,08	2,05	2,03	1,98	1,94	1,90	1,88	1,84	1,78	1,75	1,68	1,63	1,62	60
65	2,13	2,09	2,06	2,03	2,00	1,95	1,91	1,88	1,85	1,81	1,75	1,72	1,65	1,60	1,59	65
70	2,11	2,07	2,03	2,01	1,98	1,93	1,89	1,85	1,83	1,78	1,73	1,70	1,62	1,57	1,56	70
80	2,07	2,03	2,00	1,97	1,94	1,89	1,85	1,82	1,79	1,75	1,69	1,65	1,58	1,53	1,51	80
90	2,04	2,00	1,97	1,94	1,92	1,86	1,82	1,79	1,76	1,72	1,66	1,62	1,55	1,49	1,48	90
100	2,02	1,98	1,95	1,92	1,89	1,84	1,80	1,76	1,74	1,69	1,63	1,60	1,52	1,47	1,45	100
120	1,99	1,95	1,92	1,89	1,86	1,81	1,76	1,73	1,70	1,66	1,60	1,56	1,48	1,42	1,40	120
150	1,96	1,92	1,88	1,85	1,83	1,77	1,73	1,69	1,66	1,62	1,56	1,52	1,43	1,38	1,35	150
180	1,94	1,90	1,86	1,83	1,81	1,75	1,71	1,67	1,64	1,60	1,53	1,49	1,41	1,34	1,32	180
200	1,93	1,89	1,85	1,82	1,79	1,74	1,69	1,66	1,63	1,58	1,52	1,48	1,39	1,33	1,30	200
250	1,91	1,87	1,83	1,80	1,77	1,72	1,67	1,64	1,61	1,56	1,50	1,46	1,36	1,30	1,27	250
300	1,89	1,85	1,82	1,79	1,76	1,70	1,66	1,62	1,59	1,55	1,48	1,44	1,35	1,28	1,25	300
350	1,88	1,84	1,81	1,78	1,75	1,70	1,65	1,61	1,58	1,54	1,47	1,43	1,33	1,26	1,23	350
400	1,88	1,84	1,80	1,77	1,75	1,69	1,64	1,61	1,58	1,53	1,46	1,42	1,32	1,25	1,22	400
500	1,87	1,83	1,79	1,76	1,74	1,68	1,63	1,60	1,57	1,52	1,45	1,41	1,31	1,23	1,20	500
600	1,86	1,82	1,79	1,76	1,73	1,67	1,63	1,59	1,56	1,51	1,44	1,40	1,30	1,22	1,19	600
700	1,86	1,82	1,78	1,75	1,72	1,67	1,62	1,58	1,55	1,50	1,44	1,39	1,29	1,21	1,18	700
800	1,85	1,81	1,78	1,75	1,72	1,66	1,62	1,58	1,55	1,50	1,43	1,39	1,29	1,20	1,17	800
999	1,85	1,81	1,77	1,74	1,72	1,66	1,61	1,58	1,54	1,50	1,43	1,38	1,28	1,19	1,16	999

Tablica 7 – część I
Wartości krytyczne rozkładu F-Snedecora dla poziomu istotności 0,05 $F_{0,05;k_1,k_2}$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	$k_1 \backslash k_2$
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	245,36	246,46	247,32	248,01	1
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,44	19,45	2
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,69	8,67	8,66	3
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,84	5,82	5,80	4
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,60	4,58	4,56	5
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,92	3,90	3,87	6
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,49	3,47	3,44	7
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,20	3,17	3,15	8
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	2,99	2,96	2,94	9
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,83	2,80	2,77	10
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,70	2,67	2,65	11
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,60	2,57	2,54	12
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,51	2,48	2,46	13
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,44	2,41	2,39	14
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,38	2,35	2,33	15
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,33	2,30	2,28	16
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,29	2,26	2,23	17
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,25	2,22	2,19	18
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,21	2,18	2,16	19
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,18	2,15	2,12	20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,16	2,12	2,10	21
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,13	2,10	2,07	22
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,11	2,08	2,05	23
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,09	2,05	2,03	24
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,07	2,04	2,01	25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,05	2,02	1,99	26
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,04	2,00	1,97	27
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,02	1,99	1,96	28
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,01	1,97	1,94	29
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	1,99	1,96	1,93	30

Tablica 7 – część I cd.

32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,07	2,01	1,97	1,94	1,91	32
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,05	1,99	1,95	1,92	1,89	34
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,03	1,98	1,93	1,90	1,87	36
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,02	1,96	1,92	1,88	1,85	38
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,90	1,87	1,84	40
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	1,99	1,93	1,89	1,86	1,83	42
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	1,98	1,92	1,88	1,84	1,81	44
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	1,97	1,91	1,87	1,83	1,80	46
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,96	1,90	1,86	1,82	1,79	48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95	1,89	1,85	1,81	1,78	50
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,93	1,88	1,83	1,79	1,76	55
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,82	1,78	1,75	60
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,90	1,85	1,80	1,76	1,73	65
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,89	1,84	1,79	1,75	1,72	70
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,82	1,77	1,73	1,70	80
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,86	1,80	1,76	1,72	1,69	90
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,75	1,71	1,68	100
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,78	1,73	1,69	1,66	120
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,82	1,76	1,71	1,67	1,64	150
180	3,89	3,05	2,65	2,42	2,26	2,15	2,06	1,99	1,93	1,88	1,81	1,75	1,70	1,66	1,63	180
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,74	1,69	1,66	1,62	200
250	3,88	3,03	2,64	2,41	2,25	2,13	2,05	1,98	1,92	1,87	1,79	1,73	1,68	1,65	1,61	250
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,78	1,72	1,68	1,64	1,61	300
350	3,87	3,02	2,63	2,40	2,24	2,12	2,04	1,96	1,91	1,86	1,78	1,72	1,67	1,63	1,60	350
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,78	1,72	1,67	1,63	1,60	400
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,77	1,71	1,66	1,62	1,59	500
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,77	1,71	1,66	1,62	1,59	600
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,77	1,71	1,66	1,62	1,59	700
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,66	1,62	1,58	800
999	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,65	1,61	1,58	999

Tablica 7 – część II
Wartości krytyczne rozkładu F-Snedecora dla poziomu istotności 0,05 $F_{0,05;k_1,k_2}$

$k_2 \backslash k_1$	22	24	26	28	30	35	40	45	50	60	80	100	200	500	999	$k_1 \backslash k_2$
1	248,58	249,05	249,45	249,80	250,10	250,69	251,14	251,49	251,77	252,20	252,72	253,04	253,68	254,06	254,19	1
2	19,45	19,45	19,46	19,46	19,46	19,47	19,47	19,47	19,48	19,48	19,48	19,49	19,49	19,49	19,49	2
3	8,65	8,64	8,63	8,62	8,62	8,60	8,59	8,59	8,58	8,57	8,56	8,55	8,54	8,53	8,53	3
4	5,79	5,77	5,76	5,75	5,75	5,73	5,72	5,71	5,70	5,69	5,67	5,66	5,65	5,64	5,63	4
5	4,54	4,53	4,52	4,50	4,50	4,48	4,46	4,45	4,44	4,43	4,41	4,41	4,39	4,37	4,37	5
6	3,86	3,84	3,83	3,82	3,81	3,79	3,77	3,76	3,75	3,74	3,72	3,71	3,69	3,68	3,67	6
7	3,43	3,41	3,40	3,39	3,38	3,36	3,34	3,33	3,32	3,30	3,29	3,27	3,25	3,24	3,23	7
8	3,13	3,12	3,10	3,09	3,08	3,06	3,04	3,03	3,02	3,01	2,99	2,97	2,95	2,94	2,93	8
9	2,92	2,90	2,89	2,87	2,86	2,84	2,83	2,81	2,80	2,79	2,77	2,76	2,73	2,72	2,71	9
10	2,75	2,74	2,72	2,71	2,70	2,68	2,66	2,65	2,64	2,62	2,60	2,59	2,56	2,55	2,54	10
11	2,63	2,61	2,59	2,58	2,57	2,55	2,53	2,52	2,51	2,49	2,47	2,46	2,43	2,42	2,41	11
12	2,52	2,51	2,49	2,48	2,47	2,44	2,43	2,41	2,40	2,38	2,36	2,35	2,32	2,31	2,30	12
13	2,44	2,42	2,41	2,39	2,38	2,36	2,34	2,33	2,31	2,30	2,27	2,26	2,23	2,22	2,21	13
14	2,37	2,35	2,33	2,32	2,31	2,28	2,27	2,25	2,24	2,22	2,20	2,19	2,16	2,14	2,14	14
15	2,31	2,29	2,27	2,26	2,25	2,22	2,20	2,19	2,18	2,16	2,14	2,12	2,10	2,08	2,07	15
16	2,25	2,24	2,22	2,21	2,19	2,17	2,15	2,14	2,12	2,11	2,08	2,07	2,04	2,02	2,02	16
17	2,21	2,19	2,17	2,16	2,15	2,12	2,10	2,09	2,08	2,06	2,03	2,02	1,99	1,97	1,97	17
18	2,17	2,15	2,13	2,12	2,11	2,08	2,06	2,05	2,04	2,02	1,99	1,98	1,95	1,93	1,92	18
19	2,13	2,11	2,10	2,08	2,07	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,91	1,89	1,88	19
20	2,10	2,08	2,07	2,05	2,04	2,01	1,99	1,98	1,97	1,95	1,92	1,91	1,88	1,86	1,85	20
21	2,07	2,05	2,04	2,02	2,01	1,98	1,96	1,95	1,94	1,92	1,89	1,88	1,84	1,82	1,82	21
22	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,92	1,91	1,89	1,86	1,85	1,82	1,80	1,79	22
23	2,02	2,01	1,99	1,97	1,96	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,79	1,77	1,76	23
24	2,00	1,98	1,97	1,95	1,94	1,91	1,89	1,88	1,86	1,84	1,82	1,80	1,77	1,75	1,74	24
25	1,98	1,96	1,95	1,93	1,92	1,89	1,87	1,86	1,84	1,82	1,80	1,78	1,75	1,73	1,72	25
26	1,97	1,95	1,93	1,91	1,90	1,87	1,85	1,84	1,82	1,80	1,78	1,76	1,73	1,71	1,70	26
27	1,95	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,81	1,79	1,76	1,74	1,71	1,69	1,68	27
28	1,93	1,91	1,90	1,88	1,87	1,84	1,82	1,80	1,79	1,77	1,74	1,73	1,69	1,67	1,66	28
29	1,92	1,90	1,88	1,87	1,85	1,83	1,81	1,79	1,77	1,75	1,73	1,71	1,67	1,65	1,65	29
30	1,91	1,89	1,87	1,85	1,84	1,81	1,79	1,77	1,76	1,74	1,71	1,70	1,66	1,64	1,63	30

Tablica 7 – część II cd.

32	1,88	1,86	1,85	1,83	1,82	1,79	1,77	1,75	1,74	1,71	1,69	1,67	1,63	1,61	1,60	32
34	1,86	1,84	1,82	1,81	1,80	1,77	1,75	1,73	1,71	1,69	1,66	1,65	1,61	1,59	1,58	34
36	1,85	1,82	1,81	1,79	1,78	1,75	1,73	1,71	1,69	1,67	1,64	1,62	1,59	1,56	1,56	36
38	1,83	1,81	1,79	1,77	1,76	1,73	1,71	1,69	1,68	1,65	1,62	1,61	1,57	1,54	1,54	38
40	1,81	1,79	1,77	1,76	1,74	1,72	1,69	1,67	1,66	1,64	1,61	1,59	1,55	1,53	1,52	40
42	1,80	1,78	1,76	1,74	1,73	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,53	1,51	1,50	42
44	1,79	1,77	1,75	1,73	1,72	1,69	1,67	1,65	1,63	1,61	1,58	1,56	1,52	1,49	1,49	44
46	1,78	1,76	1,74	1,72	1,71	1,68	1,65	1,64	1,62	1,60	1,57	1,55	1,51	1,48	1,47	46
48	1,77	1,75	1,73	1,71	1,70	1,67	1,64	1,62	1,61	1,59	1,56	1,54	1,49	1,47	1,46	48
50	1,76	1,74	1,72	1,70	1,69	1,66	1,63	1,61	1,60	1,58	1,54	1,52	1,48	1,46	1,45	50
55	1,74	1,72	1,70	1,68	1,67	1,64	1,61	1,59	1,58	1,55	1,52	1,50	1,46	1,43	1,42	55
60	1,72	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,56	1,53	1,50	1,48	1,44	1,41	1,40	60
65	1,71	1,69	1,67	1,65	1,63	1,60	1,58	1,56	1,54	1,52	1,49	1,46	1,42	1,39	1,38	65
70	1,70	1,67	1,65	1,64	1,62	1,59	1,57	1,55	1,53	1,50	1,47	1,45	1,40	1,37	1,36	70
80	1,68	1,65	1,63	1,62	1,60	1,57	1,54	1,52	1,51	1,48	1,45	1,43	1,38	1,35	1,34	80
90	1,66	1,64	1,62	1,60	1,59	1,55	1,53	1,51	1,49	1,46	1,43	1,41	1,36	1,33	1,31	90
100	1,65	1,63	1,61	1,59	1,57	1,54	1,52	1,49	1,48	1,45	1,41	1,39	1,34	1,31	1,30	100
120	1,63	1,61	1,59	1,57	1,55	1,52	1,50	1,47	1,46	1,43	1,39	1,37	1,32	1,28	1,27	120
150	1,61	1,59	1,57	1,55	1,54	1,50	1,48	1,45	1,44	1,41	1,37	1,34	1,29	1,25	1,24	150
180	1,60	1,58	1,56	1,54	1,52	1,49	1,46	1,44	1,42	1,39	1,35	1,33	1,27	1,23	1,22	180
200	1,60	1,57	1,55	1,53	1,52	1,48	1,46	1,43	1,41	1,39	1,35	1,32	1,26	1,22	1,21	200
250	1,58	1,56	1,54	1,52	1,50	1,47	1,44	1,42	1,40	1,37	1,33	1,31	1,25	1,20	1,18	250
300	1,58	1,55	1,53	1,51	1,50	1,46	1,43	1,41	1,39	1,36	1,32	1,30	1,23	1,19	1,17	300
350	1,57	1,55	1,53	1,51	1,49	1,46	1,43	1,41	1,39	1,36	1,32	1,29	1,23	1,18	1,16	350
400	1,57	1,54	1,52	1,50	1,49	1,45	1,42	1,40	1,38	1,35	1,31	1,28	1,22	1,17	1,15	400
500	1,56	1,54	1,52	1,50	1,48	1,45	1,42	1,40	1,38	1,35	1,30	1,28	1,21	1,16	1,14	500
600	1,56	1,54	1,51	1,50	1,48	1,44	1,41	1,39	1,37	1,34	1,30	1,27	1,20	1,15	1,13	600
700	1,56	1,53	1,51	1,49	1,48	1,44	1,41	1,39	1,37	1,34	1,29	1,27	1,20	1,15	1,12	700
800	1,56	1,53	1,51	1,49	1,47	1,44	1,41	1,39	1,37	1,34	1,29	1,26	1,20	1,14	1,12	800
999	1,55	1,53	1,51	1,49	1,47	1,43	1,41	1,38	1,36	1,33	1,29	1,26	1,19	1,13	1,11	999

Tablica 8
Wartości krytyczne rozkładu F-Snedecora dla poziomu istotności 0,10 $F_{0,10;k_1,k_2}$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	15	20	50	100	200	500	$k_1 \backslash k_2$
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	61,22	61,74	62,69	63,01	63,17	63,26	1
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,42	9,44	9,47	9,48	9,49	9,49	2
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,20	5,18	5,15	5,14	5,14	5,14	3
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,87	3,84	3,80	3,78	3,77	3,76	4
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,24	3,21	3,15	3,13	3,12	3,11	5
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,87	2,84	2,77	2,75	2,73	2,73	6
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,63	2,59	2,52	2,50	2,48	2,48	7
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,46	2,42	2,35	2,32	2,31	2,30	8
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,34	2,30	2,22	2,19	2,17	2,17	9
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,24	2,20	2,12	2,09	2,07	2,06	10
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,17	2,12	2,04	2,01	1,99	1,98	11
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,10	2,06	1,97	1,94	1,92	1,91	12
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,05	2,01	1,92	1,88	1,86	1,85	13
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,01	1,96	1,87	1,83	1,82	1,80	14
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	1,97	1,92	1,83	1,79	1,77	1,76	15
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,94	1,89	1,79	1,76	1,74	1,73	16
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,91	1,86	1,76	1,73	1,71	1,69	17
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,89	1,84	1,74	1,70	1,68	1,67	18
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,86	1,81	1,71	1,67	1,65	1,64	19
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,84	1,79	1,69	1,65	1,63	1,62	20
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,83	1,78	1,67	1,63	1,61	1,60	21
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,81	1,76	1,65	1,61	1,59	1,58	22
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,80	1,74	1,64	1,59	1,57	1,56	23
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,78	1,73	1,62	1,58	1,56	1,54	24
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,77	1,72	1,61	1,56	1,54	1,53	25
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,76	1,71	1,59	1,55	1,53	1,51	26
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,75	1,70	1,58	1,54	1,52	1,50	27
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,74	1,69	1,57	1,53	1,50	1,49	28
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,73	1,68	1,56	1,52	1,49	1,48	29
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,72	1,67	1,55	1,51	1,48	1,47	30

Tablica 8 cd.

32	2,87	2,48	2,26	2,13	2,04	1,97	1,91	1,87	1,83	1,81	1,71	1,65	1,53	1,49	1,46	1,45	32
34	2,86	2,47	2,25	2,12	2,02	1,96	1,90	1,86	1,82	1,79	1,69	1,64	1,52	1,47	1,45	1,43	34
36	2,85	2,46	2,24	2,11	2,01	1,94	1,89	1,85	1,81	1,78	1,68	1,63	1,51	1,46	1,43	1,42	36
38	2,84	2,45	2,23	2,10	2,01	1,94	1,88	1,84	1,80	1,77	1,67	1,61	1,49	1,45	1,42	1,40	38
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,66	1,61	1,48	1,43	1,41	1,39	40
42	2,83	2,43	2,22	2,08	1,99	1,92	1,86	1,82	1,78	1,75	1,65	1,60	1,47	1,42	1,40	1,38	42
44	2,82	2,43	2,21	2,08	1,98	1,91	1,86	1,81	1,78	1,75	1,65	1,59	1,46	1,41	1,39	1,37	44
46	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	1,74	1,64	1,58	1,46	1,40	1,38	1,36	46
48	2,81	2,42	2,20	2,07	1,97	1,90	1,85	1,80	1,77	1,73	1,63	1,57	1,45	1,40	1,37	1,35	48
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,63	1,57	1,44	1,39	1,36	1,34	50
55	2,80	2,40	2,19	2,05	1,95	1,88	1,83	1,78	1,75	1,72	1,61	1,55	1,43	1,37	1,34	1,32	55
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,60	1,54	1,41	1,36	1,33	1,31	60
65	2,78	2,39	2,17	2,03	1,94	1,87	1,81	1,77	1,73	1,70	1,59	1,53	1,40	1,35	1,31	1,29	65
70	2,78	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72	1,69	1,59	1,53	1,39	1,34	1,30	1,28	70
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68	1,57	1,51	1,38	1,32	1,28	1,26	80
90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67	1,56	1,50	1,36	1,30	1,27	1,25	90
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66	1,56	1,49	1,35	1,29	1,26	1,23	100
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,55	1,48	1,34	1,28	1,24	1,21	120
150	2,74	2,34	2,12	1,98	1,89	1,81	1,76	1,71	1,67	1,64	1,53	1,47	1,33	1,26	1,22	1,19	150
180	2,73	2,33	2,11	1,98	1,88	1,81	1,75	1,70	1,67	1,63	1,53	1,46	1,32	1,25	1,21	1,18	180
200	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66	1,63	1,52	1,46	1,31	1,24	1,20	1,17	200
250	2,73	2,32	2,11	1,97	1,87	1,80	1,74	1,69	1,66	1,62	1,51	1,45	1,30	1,23	1,19	1,15	250
300	2,72	2,32	2,10	1,96	1,87	1,79	1,74	1,69	1,65	1,62	1,51	1,45	1,29	1,22	1,18	1,14	300
350	2,72	2,32	2,10	1,96	1,86	1,79	1,73	1,69	1,65	1,62	1,51	1,44	1,29	1,22	1,17	1,14	350
400	2,72	2,32	2,10	1,96	1,86	1,79	1,73	1,69	1,65	1,61	1,50	1,44	1,29	1,21	1,17	1,13	400
500	2,72	2,31	2,09	1,96	1,86	1,79	1,73	1,68	1,64	1,61	1,50	1,44	1,28	1,21	1,16	1,12	500
600	2,71	2,31	2,09	1,95	1,86	1,78	1,73	1,68	1,64	1,61	1,50	1,43	1,28	1,20	1,16	1,12	600
700	2,71	2,31	2,09	1,95	1,86	1,78	1,73	1,68	1,64	1,61	1,50	1,43	1,28	1,20	1,15	1,11	700
800	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61	1,50	1,43	1,28	1,20	1,15	1,11	800
999	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61	1,49	1,43	1,27	1,20	1,15	1,10	999

Tablica 9 – część I

Granice przedziału ufności dla frakcji na poziomie ufności 0,95 przy k elementach posiadających wyróżnioną cechę w próbie n -elementowej

$n-k$ k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$n-k$ k
0	– –	0,975 0,000	0,842 0,000	0,708 0,000	0,602 0,000	0,522 0,000	0,459 0,000	0,410 0,000	0,369 0,000	0,336 0,000	0,308 0,000	0,285 0,000	0,265 0,000	0,247 0,000	0,232 0,000	0
1	1,000 0,025	0,987 0,013	0,906 0,008	0,806 0,006	0,716 0,005	0,641 0,004	0,579 0,004	0,527 0,003	0,483 0,003	0,445 0,003	0,413 0,002	0,385 0,002	0,360 0,002	0,339 0,002	0,319 0,002	1
2	1,000 0,158	0,992 0,094	0,932 0,068	0,853 0,053	0,777 0,043	0,710 0,037	0,651 0,032	0,600 0,028	0,556 0,025	0,518 0,023	0,484 0,021	0,454 0,019	0,428 0,018	0,405 0,017	0,383 0,016	2
3	1,000 0,292	0,994 0,194	0,947 0,147	0,882 0,118	0,816 0,099	0,755 0,085	0,701 0,075	0,652 0,067	0,610 0,060	0,572 0,055	0,538 0,050	0,508 0,047	0,481 0,043	0,456 0,040	0,434 0,038	3
4	1,000 0,398	0,995 0,284	0,957 0,223	0,901 0,184	0,843 0,157	0,788 0,137	0,738 0,122	0,692 0,109	0,651 0,099	0,614 0,091	0,581 0,084	0,551 0,078	0,524 0,073	0,499 0,068	0,476 0,064	4
5	1,000 0,478	0,996 0,359	0,963 0,290	0,915 0,245	0,863 0,212	0,813 0,187	0,766 0,167	0,723 0,151	0,684 0,139	0,649 0,128	0,616 0,118	0,587 0,110	0,560 0,103	0,535 0,097	0,512 0,091	5
6	1,000 0,541	0,996 0,421	0,968 0,349	0,925 0,299	0,878 0,262	0,833 0,234	0,789 0,211	0,749 0,192	0,711 0,177	0,677 0,163	0,646 0,152	0,617 0,142	0,590 0,133	0,565 0,126	0,543 0,119	6
7	1,000 0,590	0,997 0,473	0,972 0,400	0,933 0,348	0,891 0,308	0,849 0,277	0,808 0,251	0,770 0,230	0,734 0,213	0,701 0,198	0,671 0,184	0,643 0,173	0,616 0,163	0,592 0,154	0,570 0,146	7
8	1,000 0,631	0,997 0,517	0,975 0,444	0,940 0,390	0,901 0,349	0,861 0,316	0,823 0,289	0,787 0,266	0,753 0,247	0,722 0,230	0,692 0,215	0,665 0,203	0,639 0,191	0,616 0,181	0,593 0,172	8
9	1,000 0,664	0,997 0,555	0,977 0,482	0,945 0,428	0,909 0,386	0,872 0,351	0,837 0,323	0,802 0,299	0,770 0,278	0,740 0,260	0,711 0,244	0,685 0,231	0,660 0,218	0,636 0,207	0,615 0,197	9
10	1,000 0,692	0,998 0,587	0,979 0,516	0,950 0,462	0,916 0,419	0,882 0,384	0,848 0,354	0,816 0,329	0,785 0,308	0,756 0,289	0,728 0,272	0,702 0,257	0,678 0,244	0,655 0,232	0,634 0,221	10

Tablica 9 – część I cd.

11	1,000	0,998	0,981	0,953	0,922	0,890	0,858	0,827	0,797	0,769	0,743	0,718	0,694	0,672	0,651	11
	0,715	0,615	0,546	0,492	0,449	0,413	0,383	0,357	0,335	0,315	0,298	0,282	0,268	0,256	0,244	
12	1,000	0,998	0,982	0,957	0,927	0,897	0,867	0,837	0,809	0,782	0,756	0,732	0,709	0,687	0,666	12
	0,735	0,640	0,572	0,519	0,476	0,440	0,410	0,384	0,361	0,340	0,322	0,306	0,291	0,278	0,266	
13	1,000	0,998	0,983	0,960	0,932	0,903	0,874	0,846	0,819	0,793	0,768	0,744	0,722	0,701	0,680	13
	0,753	0,661	0,595	0,544	0,501	0,465	0,435	0,408	0,384	0,364	0,345	0,328	0,313	0,299	0,287	
14	1,000	0,998	0,984	0,962	0,936	0,909	0,881	0,854	0,828	0,803	0,779	0,756	0,734	0,713	0,694	14
	0,768	0,681	0,617	0,566	0,524	0,488	0,457	0,430	0,407	0,385	0,366	0,349	0,334	0,320	0,306	
15	1,000	0,998	0,985	0,964	0,939	0,913	0,887	0,861	0,836	0,812	0,789	0,766	0,745	0,725	0,705	15
	0,782	0,698	0,636	0,586	0,544	0,509	0,478	0,451	0,427	0,406	0,386	0,369	0,353	0,339	0,325	
16	1,000	0,999	0,986	0,966	0,943	0,918	0,893	0,868	0,844	0,820	0,798	0,776	0,755	0,736	0,717	16
	0,794	0,713	0,653	0,604	0,563	0,529	0,498	0,471	0,447	0,425	0,405	0,388	0,372	0,357	0,343	
17	1,000	0,999	0,987	0,968	0,946	0,922	0,898	0,874	0,851	0,828	0,806	0,785	0,765	0,745	0,727	17
	0,805	0,727	0,669	0,621	0,581	0,547	0,516	0,488	0,465	0,443	0,423	0,406	0,389	0,374	0,360	
18	1,000	0,999	0,988	0,970	0,948	0,925	0,902	0,879	0,857	0,835	0,814	0,793	0,773	0,755	0,736	18
	0,815	0,740	0,683	0,637	0,597	0,564	0,533	0,506	0,482	0,460	0,440	0,422	0,406	0,391	0,376	
19	1,000	0,999	0,988	0,971	0,950	0,929	0,906	0,884	0,862	0,841	0,821	0,801	0,782	0,763	0,745	19
	0,824	0,751	0,696	0,651	0,612	0,579	0,549	0,522	0,508	0,476	0,456	0,439	0,422	0,408	0,392	
20	1,000	0,999	0,989	0,972	0,953	0,932	0,910	0,889	0,868	0,847	0,827	0,808	0,789	0,771	0,753	20
	0,832	0,762	0,708	0,664	0,626	0,593	0,564	0,537	0,513	0,492	0,472	0,454	0,437	0,421	0,407	
21	1,000	0,999	0,990	0,973	0,955	0,934	0,914	0,893	0,874	0,853	0,833	0,814	0,796	0,778	0,761	21
	0,839	0,772	0,720	0,676	0,640	0,607	0,577	0,551	0,528	0,506	0,486	0,468	0,451	0,436	0,421	
22	1,000	0,999	0,990	0,975	0,956	0,937	0,917	0,897	0,877	0,858	0,839	0,820	0,803	0,785	0,768	22
	0,846	0,781	0,730	0,688	0,651	0,619	0,590	0,565	0,541	0,519	0,500	0,481	0,465	0,449	0,434	
23	1,000	0,999	0,990	0,976	0,957	0,939	0,920	0,901	0,881	0,862	0,844	0,826	0,809	0,792	0,775	23
	0,852	0,789	0,740	0,699	0,662	0,631	0,603	0,577	0,544	0,533	0,513	0,495	0,478	0,462	0,448	
24	1,000	0,999	0,991	0,976	0,960	0,942	0,923	0,904	0,885	0,867	0,849	0,831	0,814	0,798	0,782	24
	0,858	0,797	0,749	0,708	0,673	0,642	0,614	0,589	0,566	0,545	0,525	0,507	0,490	0,475	0,460	
25	1,000	0,999	0,991	0,977	0,961	0,944	0,925	0,907	0,889	0,871	0,854	0,836	0,820	0,804	0,788	25
	0,863	0,804	0,757	0,718	0,683	0,653	0,625	0,600	0,577	0,556	0,537	0,519	0,502	0,487	0,472	

Tablica 9 – część II

Granice przedziału ufności dla frakcji na poziomie ufności 0,95 przy k elementach posiadających wyróżnioną cechę w próbie n -elementowej

$\begin{matrix} n-k \\ k \end{matrix}$	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	$\begin{matrix} n-k \\ k \end{matrix}$
0	0,218 0,000	0,206 0,000	0,195 0,000	0,185 0,000	0,176 0,000	0,168 0,000	0,161 0,000	0,154 0,000	0,148 0,000	0,142 0,000	0,137 0,000	0,132 0,000	0,127 0,000	0,123 0,000	0,119 0,000	0
1	0,302 0,002	0,287 0,001	0,273 0,001	0,260 0,001	0,249 0,001	0,238 0,001	0,228 0,001	0,219 0,001	0,211 0,001	0,203 0,001	0,196 0,001	0,190 0,001	0,184 0,001	0,178 0,001	0,172 0,001	1
2	0,364 0,015	0,347 0,014	0,331 0,013	0,317 0,012	0,304 0,012	0,292 0,011	0,280 0,010	0,270 0,010	0,260 0,010	0,251 0,009	0,243 0,009	0,235 0,009	0,228 0,008	0,221 0,008	0,215 0,008	2
3	0,414 0,036	0,396 0,034	0,379 0,032	0,363 0,030	0,349 0,029	0,336 0,028	0,324 0,027	0,312 0,025	0,301 0,024	0,292 0,024	0,282 0,023	0,274 0,022	0,265 0,021	0,257 0,020	0,250 0,020	3
4	0,456 0,061	0,437 0,057	0,419 0,054	0,403 0,052	0,388 0,050	0,374 0,047	0,360 0,045	0,349 0,044	0,338 0,042	0,327 0,040	0,317 0,039	0,307 0,038	0,298 0,036	0,290 0,035	0,282 0,034	4
5	0,491 0,087	0,471 0,082	0,453 0,078	0,436 0,075	0,421 0,071	0,407 0,068	0,393 0,066	0,381 0,063	0,369 0,061	0,358 0,058	0,347 0,056	0,337 0,055	0,328 0,053	0,319 0,051	0,311 0,050	5
6	0,522 0,113	0,502 0,107	0,484 0,102	0,467 0,098	0,451 0,094	0,436 0,090	0,423 0,086	0,410 0,083	0,397 0,080	0,386 0,077	0,375 0,075	0,364 0,072	0,355 0,070	0,345 0,068	0,336 0,066	6
7	0,549 0,139	0,529 0,132	0,512 0,126	0,494 0,121	0,478 0,116	0,463 0,111	0,499 0,107	0,435 0,103	0,423 0,099	0,411 0,096	0,400 0,093	0,389 0,090	0,379 0,087	0,369 0,084	0,360 0,082	7
8	0,573 0,164	0,553 0,156	0,535 0,149	0,518 0,143	0,502 0,138	0,487 0,132	0,472 0,126	0,459 0,123	0,446 0,119	0,434 0,115	0,423 0,111	0,412 0,107	0,401 0,104	0,391 0,101	0,382 0,098	8
9	0,594 0,188	0,575 0,180	0,557 0,172	0,540 0,165	0,524 0,159	0,508 0,153	0,494 0,147	0,481 0,142	0,467 0,138	0,455 0,133	0,444 0,129	0,433 0,125	0,422 0,121	0,412 0,118	0,402 0,114	9
10	0,614 0,211	0,595 0,202	0,577 0,194	0,560 0,186	0,544 0,179	0,528 0,173	0,514 0,167	0,500 0,161	0,487 0,156	0,475 0,151	0,463 0,146	0,452 0,142	0,441 0,138	0,431 0,134	0,421 0,130	10

Tablica 9 – część II cd.

11	0,631	0,612	0,594	0,578	0,561	0,546	0,532	0,519	0,505	0,493	0,481	0,470	0,459	0,449	0,439	11
	0,234	0,224	0,215	0,207	0,199	0,192	0,186	0,180	0,174	0,169	0,164	0,159	0,154	0,150	0,146	
12	0,647	0,628	0,611	0,594	0,578	0,563	0,549	0,535	0,522	0,510	0,498	0,487	0,476	0,465	0,455	12
	0,255	0,245	0,235	0,227	0,218	0,211	0,204	0,197	0,191	0,186	0,180	0,175	0,170	0,166	0,161	
13	0,661	0,643	0,626	0,609	0,594	0,579	0,564	0,551	0,538	0,525	0,513	0,503	0,491	0,481	0,471	13
	0,275	0,264	0,255	0,245	0,237	0,229	0,222	0,215	0,208	0,202	0,196	0,191	0,186	0,181	0,176	
14	0,675	0,657	0,640	0,624	0,608	0,593	0,579	0,556	0,552	0,540	0,528	0,517	0,506	0,496	0,485	14
	0,295	0,283	0,273	0,264	0,255	0,247	0,239	0,232	0,225	0,218	0,212	0,206	0,201	0,196	0,191	
15	0,687	0,669	0,653	0,637	0,621	0,607	0,592	0,579	0,566	0,554	0,542	0,531	0,520	0,509	0,499	15
	0,313	0,302	0,291	0,281	0,272	0,263	0,255	0,248	0,240	0,234	0,227	0,221	0,216	0,210	0,205	
16	0,698	0,681	0,665	0,649	0,634	0,619	0,605	0,592	0,579	0,567	0,555	0,544	0,533	0,522	0,512	16
	0,331	0,319	0,308	0,298	0,288	0,280	0,271	0,263	0,255	0,249	0,242	0,236	0,230	0,224	0,219	
17	0,709	0,692	0,676	0,660	0,645	0,631	0,617	0,604	0,591	0,579	0,567	0,556	0,545	0,535	0,524	17
	0,347	0,335	0,324	0,314	0,304	0,295	0,286	0,278	0,270	0,263	0,256	0,250	0,244	0,238	0,232	
18	0,719	0,702	0,686	0,671	0,656	0,642	0,628	0,615	0,602	0,590	0,579	0,568	0,557	0,547	0,536	18
	0,363	0,351	0,340	0,329	0,319	0,310	0,301	0,293	0,285	0,277	0,270	0,264	0,257	0,251	0,245	
19	0,728	0,712	0,696	0,681	0,666	0,652	0,639	0,626	0,613	0,601	0,590	0,578	0,568	0,557	0,547	19
	0,379	0,366	0,355	0,344	0,334	0,324	0,315	0,307	0,298	0,291	0,284	0,277	0,270	0,264	0,258	
20	0,737	0,720	0,705	0,690	0,676	0,662	0,649	0,636	0,623	0,612	0,600	0,589	0,578	0,568	0,588	20
	0,393	0,381	0,369	0,358	0,348	0,338	0,329	0,320	0,312	0,304	0,296	0,289	0,283	0,276	0,270	
21	0,745	0,729	0,714	0,699	0,685	0,671	0,658	0,645	0,633	0,621	0,610	0,599	0,588	0,578	0,568	21
	0,408	0,395	0,383	0,372	0,361	0,351	0,342	0,333	0,325	0,317	0,309	0,302	0,295	0,288	0,282	
22	0,752	0,737	0,722	0,707	0,693	0,680	0,667	0,654	0,642	0,631	0,619	0,608	0,598	0,587	0,577	22
	0,421	0,408	0,396	0,385	0,374	0,364	0,355	0,346	0,337	0,329	0,321	0,314	0,307	0,300	0,293	
23	0,760	0,745	0,730	0,715	0,702	0,688	0,675	0,663	0,651	0,639	0,628	0,617	0,607	0,596	0,586	23
	0,434	0,421	0,409	0,398	0,387	0,377	0,367	0,358	0,349	0,341	0,333	0,325	0,318	0,311	0,304	
24	0,766	0,751	0,737	0,723	0,709	0,696	0,683	0,671	0,659	0,648	0,636	0,626	0,615	0,605	0,595	24
	0,446	0,433	0,421	0,410	0,399	0,388	0,379	0,370	0,361	0,352	0,344	0,337	0,330	0,322	0,315	

Tablica 10
Kwantyle $d_n(1 - \alpha)$ statystyki D_n Kołmogorowa dla poziomu istotności α

n	α			n	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	0,950	0,975	0,995	51	0,168	0,187	0,224
2	0,776	0,842	0,929	52	0,166	0,185	0,222
3	0,636	0,708	0,829	53	0,165	0,183	0,220
4	0,565	0,624	0,734	54	0,163	0,181	0,218
5	0,509	0,563	0,669	55	0,162	0,180	0,216
6	0,468	0,519	0,617	56	0,160	0,178	0,214
7	0,436	0,483	0,576	57	0,159	0,177	0,212
8	0,410	0,454	0,542	58	0,158	0,175	0,210
9	0,387	0,430	0,513	59	0,156	0,174	0,208
10	0,369	0,409	0,489	60	0,155	0,172	0,207
11	0,352	0,391	0,468	61	0,154	0,171	0,205
12	0,338	0,375	0,449	62	0,153	0,170	0,203
13	0,325	0,361	0,432	63	0,151	0,168	0,202
14	0,314	0,349	0,418	64	0,150	0,167	0,200
15	0,304	0,338	0,404	65	0,149	0,166	0,199
16	0,295	0,327	0,392	66	0,148	0,164	0,197
17	0,286	0,318	0,381	67	0,147	0,163	0,196
18	0,279	0,309	0,371	68	0,146	0,162	0,194
19	0,271	0,301	0,361	69	0,145	0,161	0,193
20	0,265	0,294	0,352	70	0,144	0,160	0,192
21	0,259	0,287	0,344	71	0,143	0,159	0,190
22	0,253	0,281	0,337	72	0,142	0,158	0,189
23	0,247	0,275	0,330	73	0,141	0,156	0,188
24	0,242	0,269	0,323	74	0,140	0,155	0,186
25	0,238	0,264	0,317	75	0,139	0,154	0,185
26	0,233	0,259	0,311	76	0,138	0,153	0,184
27	0,229	0,254	0,305	77	0,137	0,152	0,183
28	0,225	0,250	0,300	78	0,136	0,151	0,182
29	0,221	0,246	0,294	79	0,136	0,151	0,181
30	0,218	0,242	0,290	80	0,135	0,150	0,179
31	0,214	0,238	0,285	81	0,134	0,149	0,178
32	0,211	0,234	0,281	82	0,133	0,148	0,177
33	0,208	0,231	0,277	83	0,132	0,147	0,176
34	0,205	0,227	0,273	84	0,131	0,146	0,175
35	0,202	0,224	0,269	85	0,131	0,145	0,174

Tablica 10 cd.

n	α			n	α		
	0,10	0,05	0,01		0,10	0,05	0,01
36	0,199	0,221	0,265	86	0,130	0,144	0,173
37	0,196	0,218	0,262	87	0,129	0,144	0,172
38	0,194	0,215	0,258	88	0,128	0,143	0,171
39	0,191	0,213	0,255	89	0,128	0,142	0,170
40	0,189	0,210	0,252	90	0,127	0,141	0,169
41	0,187	0,208	0,249	91	0,126	0,140	0,168
42	0,185	0,205	0,246	92	0,126	0,140	0,168
43	0,183	0,203	0,243	93	0,125	0,139	0,167
44	0,181	0,201	0,241	94	0,124	0,138	0,166
45	0,179	0,198	0,238	95	0,124	0,137	0,165
46	0,177	0,196	0,235	96	0,123	0,137	0,164
47	0,175	0,194	0,233	97	0,122	0,136	0,163
48	0,173	0,192	0,231	98	0,122	0,135	0,162
49	0,171	0,190	0,228	99	0,121	0,135	0,162
50	0,170	0,188	0,226	100	0,121	0,134	0,161

Literatura

- [1] Aczel Amir D., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa 2000.
- [2] Burzyński J., *Rachunek prawdopodobieństwa i statystyka matematyczna dla wydziałów technicznych i przyrodniczych akademii rolniczych*, Akademia Rolnicza im. H. Kołłątaja, Kraków 1982.
- [3] Czaja J., *Modele statystyczne w informacji o terenie*, Wydawnictwa AGH, Kraków 1996.
- [4] Feller W., *An introduction to Probability Theory and Its Applications*, tom 2, Wiley, New York 1971.
- [5] Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, Warszawa 1976.
- [6] Hellwig Z., *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa 1998.
- [7] Koronacki J., Mielniczuk J., *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT, Warszawa 2001.
- [8] Krysiński W. i in., *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, PWN, Warszawa 1995.
- [9] Ombach J., *Wstęp do rachunku prawdopodobieństwa*, Wydawnictwa UJ, Kraków 1993.
- [10] Ostasiewicz S., Rusnak Z., Siedlecka U., *Statystyka – elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2005.
- [11] Pluciński A.E., *Probabilistyka*, WNT, Warszawa 2000.
- [12] Smirnow N.W., Dunin-Barkowski I.W., *Kurs rachunku prawdopodobieństwa i statystyki matematycznej dla zastosowań technicznych*, PWN, Warszawa 1969.

Skorowidz

A

- asymptotyczna efektywność 77
- nieobciążoność 77

B

- badanie kompletne 61
- błąd drugiego rodzaju 94
- losowy 125
- pierwszego rodzaju 94

C

- cechy ilościowe 62
- jakościowe 62
- charakterystyki rozkładu zmiennej losowej 21

D

- dwuwymiarowa zmienna losowa 43
- dwuwymiarowy rozkład normalny 58
- dystrybuanta zmiennej losowej 15, 16, 43

E

- eksczes 23
- elipsa stałej gęstości
- prawdopodobieństwa 59
- estymacja 76

- przedziałowa 84
- punktowa 82
- estymator 76
- dostateczny 76
- efektywny 76
- nieobciążony 76
- zgodny 76

F

- frakcja 75
- funkcja gamma 35
- prawdopodobieństwa 16
- testowa 94
- wiarygodności 80

G

- gęstość rozkładu prawdopodobieństwa 19, 48, 49

H

- hipoteza alternatywna 94
- statystyczna 94
- zerowa 94
- histogram rozkładu
- prawdopodobieństwa 16

K

kowariancja 51
krzywa Gaussa 33
kurtoza 23
kwantyl 23

L

liniowa regresja wieloraka 126

M

macierz kowariancji 52
metoda najmniejszych kwadratów 77
– największej wiarygodności 80
miara probabilistyczna 12
moment centralny 22
– – absolutny 67
– – mieszany 50
– mieszany 50
– zwykły 22
– – absolutny 67
– – mieszany 50

N

nierównomierność 67

O

obszar krytyczny 94
odchylenie przeciętne 67
– standardowe 22

P

populacja generalna 61
poziom istotności 94
– ufności 84
prawdopodobieństwo 12
– całkowite 13
– warunkowe 12

prosta regresji pierwszego rodzaju 59
– – drugiego rodzaju 55
próba losowa 61
próba statystyczna 61, 76
przedział ufności 84
przestrzeń zdarzeń elementarnych 11
– probabilistyczna 12

R

regresja 125
– wielowymiarowa 126
rozkład 68
– Bernoullego 26
– brzegowy 45, 48
– chi-kwadrat 35
– dwumianowy 26
– – ujemny 26
– dwupunktowy 26
– F-Snedecora 37
– Gaussa 33
– geometryczny 27
– jednopunktowy 26
– jednostajny 31
– normalny 32
– Pascala 26
– Poissona 27
– prawdopodobieństwa 15, 19
– prostokątny 31
– równomierny 25, 31
– Studenta 36
– średniej arytmetycznej z próby 68
– wariancji z próby 70
– warunkowy 46, 49
– wykładniczy 32
– zero-jedynkowy 26
rozsęp 63

S

sigma-algebra 11

skośność 23

standardowy błąd oceny 125

standaryzacja 34

standaryzowany rozkład normalny 34

statystyka 68

- matematyczna 61

- opisowa 61

szereg rozdzielczy 63

średnia arytmetyczna 65

- ważona 65

średni błąd resztowy 125

T

test chi-kwadrat Pearsona 113

- istotności 95

- Kołmogorowa 118

- najmocniejszy 94

- nieparametryczny 94

- parametryczny 94

- statystyczny 94

W

waga 65

wariancja 22

- resztowa 125

wartość krytyczna testu 94

- oczekiwana 21

- przeciętna 21, 50

- średnia 21

wektor losowy dwuwymiarowy 43

- wartości przeciętnych 52

weryfikacja hipotez statystycznych 76

wskaźnik struktury 75

współczynnik asymetrii 23

- korelacji liniowej 54

- – krzywoliniowej 125

- skupienia 23

- spłaszczenia 23

względna częstość 82

Z

zbiór zdarzeń elementarnych 11

zdarzenia niezależne 13

zdarzenie elementarne 11

- losowe 11

- niemożliwe 11

- pewne 11

- przeciwne 12

zmienna losowa 15

zmienna losowa ciągła 19, 48

- – dwuwymiarowa 43, 47

- – dyskretna (skokowa) 15, 43

- – objaśniająca (niezależna) 125

- – objaśniana (zależna) 125

zmienne losowe niezależne 21, 49

zmiennność 67

